



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 98060777**

**TO: Manjunath N Rao**  
**Location: 1/10a11 & 10d01**  
**Art Unit: 1652**  
**Monday, July 07, 2003**

**Case Serial Number: 086082**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**CM1-6B02**  
**Phone: 305-9203**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Rao,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

98801

**STIC-Biotech/ChemLib**

**From:** Rao, Manjunath N.  
**Sent:** Wednesday, July 02, 2003 3:15 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** Sequence search request for 10/086,082

**From:** Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10D 01  
Phone: 306-5681

**Date:** 7-2-03

Please search the following as soon as possible for application with serial number **10/086,082**

1. **SEQ ID NO:1**, against all commercial nucleic acid databases including issued patents database and pending application database and provide a print of **all** results.
2. **SEQ ID NO: 2** against all commercial protein databases including issued patents database and pending application database and provide a print of **all** results.

If you have any questions please call me at the above phone number.

Thanks

**Manjunath N. Rao, Ph.D.**  
**Biotechnology Patent Examiner**  
**Art Unit 1652, Room 10A11**  
**Mail Box in 10D01**  
**Crystal Mall 1, USPTO.**

Manjunath N. Rao  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall 1, USPTO

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 7/3/03  
Searcher Prep/Review: [Signature]  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: 1  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 9300  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

STIC-Biotech/ChemLib

98001

ORFE

From: Rao, Manjunath N.  
Sent: Wednesday, July 02, 2003 3:15 PM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 10/086,082

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10D 01  
Phone: 306-5681

Date: 7-2-03

Please search the following as soon as possible for application with serial number **10/086,082**

1. SEQ ID NO:1, against all commercial nucleic acid databases including issued patents database and pending application database and provide a print of all results.
2. SEQ ID NO: 2 against all commercial protein databases including issued patents database and pending application database and provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Mail Box in 10D01  
Crystal Mall 1, USPTO.

Edward Hall  
Technical Info. Specialist  
STIC/Biotech  
CM1 6B02 Tel: 305-9205

Point of Contact:  
Linda Basker  
Library Technician  
CM1 1C19 Tel: 308-4500

1-1442 na  
2-328 aa

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 7/13/03  
Searcher Prep/Review: 7/13/03  
Clerical: \_\_\_\_\_  
Online time: 6

TYPE OF SEARCH: \_\_\_\_\_  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 03/02  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



# STIC SEARCH RESULTS

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ+Desk





Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

***Because they contain data that is confidential, the results of Pending database searches should not be left in the case .***

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Ran on: July 3, 2003, 15:45:52 ; Search time 25 seconds  
(without alignments)  
371.181 Million cell updates/sec

Title: us-10-086-082-2

Perfect score: 1748

Sequence: 1 MKWLEES:NAKRGVAGRKPE.....GNVDPKTYVGAMLNKMLIV 328

Scoring table:

BLOSUM62

Gapex: 10.0 , Gapext: 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	177.5	10.2	575	US-08-922-865-2	Sequence 2, Appli
2	177.5	10.2	575	US-08-922-865-2	Sequence 2, Appli
3	91.5	5.2	1848	US-08-922-865-2	Sequence 5, Appli
4	91.5	5.2	1848	PCT-US95-10661A-6	Sequence 6, Appli
5	82.5	4.7	500	US-08-926-842R-64	Sequence 54, Appli
6	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appli
7	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appli
8	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appli
9	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appli
10	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appli
11	82.5	4.7	857	US-08-926-842R-64	Sequence 54, Appli
12	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appli
13	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appli
14	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appli
15	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appli
16	81.5	4.7	1170	US-08-926-842R-64	Sequence 54, Appli
17	80.5	4.6	1342	US-08-926-842R-64	Sequence 54, Appli
18	80.5	4.6	1342	US-08-926-842R-64	Sequence 54, Appli
19	79.5	4.5	1544	US-08-926-842R-64	Sequence 54, Appli
20	79.5	4.5	1544	US-08-926-842R-64	Sequence 54, Appli
21	79.5	4.5	1544	US-08-926-842R-64	Sequence 54, Appli
22	79.5	4.5	1544	US-08-926-842R-64	Sequence 54, Appli
23	79.5	4.5	1544	US-08-926-842R-64	Sequence 54, Appli
24	79.5	4.5	1544	US-08-926-842R-64	Sequence 54, Appli
25	79.5	4.5	1544	US-08-926-842R-64	Sequence 54, Appli
26	78.5	4.5	375	US-08-926-842R-64	Sequence 54, Appli
27	78.5	4.5	375	US-08-926-842R-64	Sequence 54, Appli

Patent No. 5268290

## ALIGNMENTS

RESULT 1

US-08-922-865-2

Sequence 2, Application: US/05922865

Patent No. 6050616

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: MICROORGANISM, LACTAMASE ENZYME OBTAINED

NUMBER OF SEQUENCES: 2

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/922,865

FILING DATE:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 575 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: Protein

US-08-922-865-2

Query Match 10.2% Score 177.5 DB 3; Length 575;

Best Local Similarity 24.0%; Pred. No. 3e-11;

Matches 59; Conservative 33; Mismatches 85; Indels 101; Gaps 12;

Qy	39	PYSTP-----VLTEPGDRIVUTRAFEAGI---NSEQDIPS-GILKHPFLN 82	
Db	44	PYCNQVHNRHFFCTPMMAVVEPCAEPKLTIDWTGCAIKNDUSAEIVRDVLSVWFL 102	
Qy	83	PONGPIMVNGAEKGVCLAV-----VIESMIFGV-----C 112	
Db	103	--SGPVGVKGAQPGCLLVLLDISAKDSSLSWSENGFFSKONGGFLDEHFFLACKSKS:MD 160	
Qy	113	PYGCAMTPIH-----FGLTGTDTLMLNPLPEKVMKILDKSEKVVWKRHFLYKPHI 167	
Db	161	PHOMTEKSRHFGVNFACLIHFGGLICLPD-----KML-----ASWNERET ----- 202	
Qy	165	GTLSVSPS-----IDNSINSLTPNHSNGMDVFGIGWG 199	
Db	203	GLIATDPRIPGIANPATTATRMQMGQKAAAGARTVPPPHGSGND:KOLSRG 262	
Qy	200	SITYPLVRAPGSKLPIGDAHACQDGEIC-GTAVEFASITTIKVDLIK 245	
Db	253	SRVFPPVVDGAGLSVGLHFSQDGEITFWGPIEMPGVHMVKSUIK 310	

## RESULT 2

US-09-510-949-2  
Sequence 2, Application US/09510949  
Patent No. 6423522  
GENERAL INFORMATION:  
APPLICANT: St. Gene III, Joseph W.  
TITLE OF INVENTION: MICROORGANISM, LACTAMASE ENZYME OBTAINED  
FROM, AND THEIR USE  
NUMBER OF SEQUENCES: 2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/510,949  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/922,655  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-510-949-2

Query Match 10.2% Score 177.5; DB 4; Length 575;  
Best Local Similarity 24.0%; Pred. No. Re-11;  
Matches 69; Conservative 33; Mismatches 85; Indels 101; Gaps 12;  
QY 39 PYSTP-----VLTIEGDRIVCTRDFAEQA-----NSKODIPS-QLLKWFLN 82  
DB 44 PYDNFOVHRWIPDIPMAVWVEPGAEEFKLETITWTGGA:KNDASAEVDREVLSTVFEF- 122  
QY 83 PONGPIKVNAGAEKGDVLAV-----YIFSMPLPROV-----E 112  
DB 103 --SGPVGAGQAGDGLLVVLLDLDIGARDLSLWGFNGFSGKNGGFLDEHEPLAKS:WD 160  
QY 113 PYGICAMIPH-----FGLGIGLIDLAMLNDPLPEKVRMIKLSKVKVWSKHTLPYKPHI 167  
DB 161 PHGMFTKSRHPGVNPAFLHPLGLGCLPDP-----KML-----ASKNPEI----- 202  
QY 168 GILSVSP-----IDNSLTPONHGGNNDVDPDIPG 199  
DB 203 GLIATDPDRIPGLANFPNATTIAHGMOMGEARDAKAAEASARTVPRHGGNCIDKLSRG 262  
QY 200 SITPLVRAPGGLFGLGHAACGGDEIC-GTAVEFASITTIKVDLIK 245  
DB 263 SRVFFPVYVDGAGLSVGDJHFSQGDGEITFMGPIENPGWVHKVSLIK 310

## RESULT 3

US-08-296-791-6  
Sequence 6, Application US/08296791  
Patent No. 6243337  
GENERAL INFORMATION:  
APPLICANT: St. Gene III, Joseph W.  
APPLICANT: Falkow, Stanley  
TITLE OF INVENTION: Haemophilus Adherence and Penetration  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/296,791  
FILING DATE: 25-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-59941/REF./RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277295  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1848 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
US-08-296-791-6

Query Match 5.2% Score 91.5; DB 4; Length 1848;  
Best Local Similarity 19.1%; Pred. No. 1.4;  
Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 20;  
QY 35 YTGIPYTPVLTIEPGD-----RIIVCTRDAFE-GAINSE-ODIPSOLLKMFNLNPGPI 88  
DB 17 YALPYTEAALVRQNDYQIFRDFAENKGFESVGA:INVEVRDKKNOSL-----GSA 67  
QY 89 MVNAGAEKGDVLAVYTESMLPRGVDPY-----GICAMIFHFGGLTGTETLAKENP 138  
DB 68 LEPGPMIDFSEVVEVQKRIATLVNPGYVGVGVKHSVGVSEL--HEGKLNCGN-----MNG 120  
QY 139 LPEKVMK:KLSKVKVYSKRIITLP-----YKPHICTLSNSPFI 177  
DB 121 NAKSHRVSSSENFYVYVEKNNEPTENVTSFTEEDQAKRRREDYTMPRLDKFTV--EVA 178  
QY 178 SINSLTPONHGGNNDVDPDIPGGSITP-LVRAFGGRFLFGHAACGGDEICGTAVEFAS 236  
DB 179 P-EASTANNKGEYNSD-----KYPAFVRLSGSTQFVYKGS-----RYQL 220  
QY 237 TTTIK-----VQIKKNQKLSNWRME-----NAENIMSISARPLE---DA 273  
DB 222 ILTEKDKGGLIRNWDGGDLVLGNAYTYGIAGIPYKVNHNHNSLIFGNSKEERSDF 280  
QY 274 TRAYRPD-----LIY-----WLV-----EDF-----GFEDTAYMLDSQ 302  
DB 281 KGLISQDPLTNVAVLSGSSPLPYVDREKCKKWLFGSYDFWAGYNNKSKWENNY----- 335  
QY 303 COKVRLGNWDFKTYTVGAMLNKN 325  
DB 336 --KHEFAEKLYQOVSAGSLIGSN 356

## RESULT 4

PCI-US95-10661A-6  
Sequence 6, Application PC/TUS9510661A  
GENERAL INFORMATION:  
APPLICANT: Washington University, et al.  
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIORITY INFORMATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: PP-59941/RET
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-6

Query Match      5.2%   Score 91.5; DB 5; Length 1848;
Best Local Similarity 19.1%; Pred. No. 1.4;
Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 20;

QY 35 VTIGYSTPVLTIEPGD-----RIIVDTDAPE-GATNSE-QDIPSOJKKXFFLNPNNGT #8
DB 17 YALTYITAEALVRDDVDYQIFRDEAKNKGFSGVATNVEVDKNQSJ-----GSA 67
QY 89 MVNKAEGDVLAVTESMLPRGVDPY-----GICAMPIREGGTLTGTLTAMNDP 138
DB 68 LPLGEMIDESVVDVKRIATLVNPQYVGKVHYSNCVSSEL--HFGNLGN-----MNG 120
QY 139 LPKYRMKLDSERVYNKRHLCP-----YKPHIGTSLVSPEID 177
DB 121 NAKSHRDVSSSENRYTVYTKRNFTPIENTVSTKGEQCAOKREDIYYPRLDKFYV---EVA 178
QY 178 S-TNSLTPDHGGMNDVPDIPGSIITYP-LVRAPGGRFLFGDAHACQGDSE-CQTAVEFAS 236
DB 179 PIEASTANNKEGYNNSD-----KYPAPFRIGSOTQFIYKKGS-----RYQL 220
QY 237 ITLIK---VOLIKNQLSHPME-----NAFLMISGARPLE---DA 278
DB 221 ILETCKOGNLRNWDVGGONLELVGNAYTYGIAGTQPKYKNHNENGLGFSNKEETSEP 280
QY 274 TRAYRD-----LIY-----WLV---EDF-----GFECWAYMLLSQ 302
DB 281 KGLSQDLPTNVAVLGDSGPLFVYDREKQKWLFGLSYDFWAGYKKSXCWEWNIY---- 315
QY 303 CCKVRLGNMDPKYTGVAGYLKN 325
DB 336 --KHFAEKIYOOSAGSLRSN 356

RESULT 5
US-08-926-842B-64
Sequence 64, Application US/08026842B
Patent No. 6030807
GENERAL INFORMATION:
APPLICANT: Sa-No. 6030807,eira, Isabel
APPLICANT: de Lencastre, Herminia
TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIORITY INFORMATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: PP-59941/RET
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-6

Query Match      5.2%   Score 91.5; DB 5; Length 1848;
Best Local Similarity 19.1%; Pred. No. 1.4;
Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 20;

QY 35 VTIGYSTPVLTIEPGD-----RIIVDTDAPE-GATNSE-QDIPSOJKKXFFLNPNNGT #8
DB 17 YALTYITAEALVRDDVDYQIFRDEAKNKGFSGVATNVEVDKNQSJ-----GSA 67
QY 89 MVNKAEGDVLAVTESMLPRGVDPY-----GICAMPIREGGTLTGTLTAMNDP 138
DB 68 LPLGEMIDESVVDVKRIATLVNPQYVGKVHYSNCVSSEL--HFGNLGN-----MNG 120
QY 139 LPKYRMKLDSERVYNKRHLCP-----YKPHIGTSLVSPEID 177
DB 121 NAKSHRDVSSSENRYTVYTKRNFTPIENTVSTKGEQCAOKREDIYYPRLDKFYV---EVA 178
QY 178 S-TNSLTPDHGGMNDVPDIPGSIITYP-LVRAPGGRFLFGDAHACQGDSE-CQTAVEFAS 236
DB 179 PIEASTANNKEGYNNSD-----KYPAPFRIGSOTQFIYKKGS-----RYQL 220
QY 237 ITLIK---VOLIKNQLSHPME-----NAFLMISGARPLE---DA 278
DB 221 ILETCKOGNLRNWDVGGONLELVGNAYTYGIAGTQPKYKNHNENGLGFSNKEETSEP 280
QY 274 TRAYRD-----LIY-----WLV---EDF-----GFECWAYMLLSQ 302
DB 281 KGLSQDLPTNVAVLGDSGPLFVYDREKQKWLFGLSYDFWAGYKKSXCWEWNIY---- 315
QY 303 CCKVRLGNMDPKYTGVAGYLKN 325
DB 336 --KHFAEKIYOOSAGSLRSN 356

RESULT 5
US-08-926-842B-64
Sequence 64, Application US/08026842B
Patent No. 6030807
GENERAL INFORMATION:
APPLICANT: Sa-No. 6030807,eira, Isabel
APPLICANT: de Lencastre, Herminia
TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/926.842B
FILING DATE: 10-SEP-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson ESS., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-089 N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1584
TELEX: 133521
INFORMATION FOR SEQ ID NO.: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORGANISM: Bacillus subtilis
FEATURE: OTHER INFORMATION: /product= "abfA"
US-08-926-842B-64

Query Match      4.7%   Score #2.5; DB 3; Length 500;
Best Local Similarity 25.4%; Pred. No. 1.7;
Matches 51; Conservative 32; Mismatches 75; Indels 43; Gaps 13;

QY 141 EKVRMIKLDSE-----KVYMSKRHTLPYKPHIGTSLVSPEIDSINSLTDRHGGMND 192
DB 2 KKARMI-VCKEYKICEVDKRIYGS-----FIEHMGR-AYVEGIVEPDHEADGEDRKD 53
QY 193 VPDIGPGSITPLVRAPGGRFLFGDAHACQGDSELCG--TAVERFAS-TTIKVLILKWKQL 250
DB 54 VQSL-IKELQVELIRYPPGNG-FLSYNMEDGVGVHNRPRMLDLAWOTIETNEVGTNEFL 11;
QY 251 SNFRMENAEIMSIG-SAPLEDATRIAYRDLI-----YW--LVEDFGFEQNDAYM 258
DB 112 SHAKAVNT-EVNMAVNLGTRSID-----AARNLVEVCNEPKGSYWSL-KRSHGIEQ---PYG 184
QY 299 ELSCGCKVPLGNKNTPKYTVG 319
DB 165 IXIWG-----LGNFMDSFWQIG 181

RESULT 6
US-08-779-113-2
Sequence 2, Application US/68779113
Patent No. 594889;
GENERAL INFORMATION:
APPLICANT: Staunton, Donald E.
APPLICANT: Harris, Edith S.
TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
TITLE OF INVENTION: Binding
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/03/779,113

FILING DATE:

CLASSIFICATION: 516

ATTORNEY/AGENT INFORMATION:

NAME: Greta E. No. 594861and

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/33773

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 857 amino acids

TYPE: amino acid

TOPOLOGY: not relevant

MOLECULE TYPE: protein

US-03-779-113-2

Query Match 4.7%: Score 82.5; DB 2; Length 857;  
Best Local Similarity 20.3%; Pred. No. 4.3;  
Matches 48; Conservative 31; Mismatches 85; Indels 73; Gaps 9;

QY 5 SESIMAKR-----GVGAGRKPYTHLTEMQKEHYHTIGPSTPVLTIPFGR-----52

DB 204 DEPCILKRLPNHHIGISLIPREVRHEHVS:KKNGHNVANSFVSINVQSEIGDARRAKVY 263

QY 53 -----LIIVTRDAFEAGINSEODIPSOILKMPFLNPONGPIMVNGAEKGV 98

DB 264 GRLSGRTFMSDFIVDTIDAGYGGIS-----LAVGPGSKVDI 302

QY 99 LAVYIESMLPSGVDPGICAMIPHEGGLTGLD--TAMLNPLPEKVKMIKLDSEKVVYS 156

DB 303 QTEDELED-----GTC-KVSYPTVPGVYIVSTKZADEHVPSPPTVKISGEG--- 348

QY 157 KRHTLPYKPHIGTSLVSFEIDSLNLTPEHNGKMWPDGPGSGITYPVLRAPGGR 213

DB 349 -----RVKESITRTSRAPSVATVGSIC-----DLNLKIPINSSDMS-AHVTSPSGRV 395

RESULT 7

US-08-583-562B-2

Sequence 2, Application US/08583562B

Patent No. 5922570

GENERAL INFORMATION:

APPLICANT: Staunton, Donald

APPLICANT: Harris, Edith

TITLE OF INVENTION: Cytoplasmic Modulators of Integrin

TITLE OF INVENTION: Binding

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sears Tower

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/583,562B

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/33038

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 858 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-583-562B-2

Query Match 4.7%: Score 82.5; DB 2; Length 858;

Best Local Similarity 20.3%; Pred. No. 4.3;

Matches 48; Conservative 31; Mismatches 85; Indels 73; Gaps 9;

QY 5 SESIMAKR-----GVGAGRKPYTHLTEMQKEHYHTIGPSTPVLTIPFGR-----52

DB 204 DEPCILKRLPNHHIGISLIPREVRHEHVS:KKNGHNVANSFVSINVQSEIGDARRAKVY 263

QY 53 -----LIIVTRDAFEAGINSEODIPSOILKMPFLNPONGPIMVNGAEKGV 98

DB 264 GRLSGRTFMSDFIVDTIDAGYGGIS-----LAVGPGSKVDI 302

QY 99 LAVYIESMLPSGVDPGICAMIPHEGGLTGLD--TAMLNPLPEKVKMIKLDSEKVVYS 156

DB 303 QTEDELED-----GTC-KVSYPTVPGVYIVSTKZADEHVPSPPTVKISGEG--- 348

QY 157 KRHTLPYKPHIGTSLVSFEIDSLNLTPEHNGKMWPDGPGSGITYPVLRAPGGR 213

DB 349 -----RVKESITRTSRAPSVATVGSIC-----DLNLKIPINSSDMS-AHVTSPSGRV 395

RESULT 8

US-09-556-877-180

Sequence 180, Application US/09556877

Patent No. 6432916

GENERAL INFORMATION:

APPLICANT: Probst, Peter

APPLICANT: Bratia, Ajay

APPLICANT: Skeiky, Yasir

APPLICANT: Fling, Steve

APPLICANT: Maisonneuve, Jeff

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

FILE REFERENCE: 210121.459C5

CURRENT APPLICATION NUMBER: US/09/556.877

CURRENT FILING DATE: 2000-04-19

NUMBER OF SEQ ID NOS: 305

SOFTWARE: FASTSEQ for Windows Version: 3.0/4.0

SEQ ID NO 180

LENGTH: 1752

TYPE: PRT

ORGANISM: Chlamydia

US-09-556-877-180

Query Match

Best Local Similarity 23.1%; Pred. No. 14;

Matches 49; Conservative 26; Mismatches 60; Indels 77; Gaps 10;

QY 40 YSTPVLTIEGCRILIVTRDAFE--GAINSEODIPSOILKMPFLNPONGPIMVNGAE--- 94

DB 271 YAGELTIADSQEVFFSINKATOGGAIFAEDV-----SFENITSLKVATNGAEKGS 323

QY 95 -----KGDVL-----AVYIESMLPRGVDPYGCAMIPHEGGLTGLD 130

DB 324 CALYAKHLSIQSSKOSLNSNYSKGGGALVE-----SGINPQD 364

QY 131 LTAMNGLPEKVKMIKLDSEKVVYWKRHITLPYKPHIGTSLVSFEIDSLNLTPEHNGN 190

DB 365 L-----EELF-IKYNKAGTETKKTIT---PSLKAQAGAGNADAWASSSSVQSSGA 419

QY 191 MCVPDIGPGS!-----TYPVLRAPGGRLF 213

DB 412 TTVSDSSGSSGSDSTSTSTVP-VIAGGGLY 442

RESULT 9

US-09-620-412C-180  
 : Sequence 180, Application US/09/620-412C  
 : Patent No. 6442354  
 : GENERAL INFORMATION:  
 : APPLICANT: Steven P. Fling  
 : TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
 : FILE REFERENCE: 210121.469C7  
 : CURRENT APPLICATION NUMBER: US/09/620.412C  
 : CURRENT FILING DATE: 2000-07-20  
 : NUMBER OF SEQ ID NOS: 363  
 : SOFTWARE: FastSeq for Windows Version 3.0/4.0  
 : SEQ ID NO 180  
 : LENGTH: 1752  
 : TYPE: PRT  
 : ORGANISM: Chlamydia  
 : US-09-620-412C-180

Query Match 4.7% Score 82.5; DB 4; Length 1752;  
 Best Local Similarity 23.1%; Pred. No. 14;  
 Matches 45; Conservative 26; Mismatches 60; Indels 77; Gaps 10;  
 QY 40 YSTPVLTIERGDIIVDTDAFE--GAINSEDDIPSOILKMFPLNPQNGPINKVNGAE--- 94  
 DB 271 YAKGDLIIADSQEVLFSINKATKDGGAFAEKDV-----SFENITSLKVTNGAEKGG 323  
 QY 95 -----KGDVL-----AVYIESMLPRGVDPYGTICAMIRIFUGLGTGD 130  
 DB 324 GAIYAKGDLISQSKOSLNSNSKGGGALAYE-----GGTNEFD 364  
 QY 131 ITANLNPPLPEKVMKLDSEKVKYKSRHTLPYKPHLGTLSVSEIUSNSLTPONHGN 190  
 DB 365 L-----EIR-IRYNKAGTFETKKTPL---PSLKACAGNADAWASSPQNGSA 411  
 QY 191 MVDVDPGCSI-----TYPVLRAPGSKLF 214  
 DB 412 TVVSDSDSSSGSUSDTSEIVF-VTAKGGGLY 442

RESULT 10  
 US-09-177-165A-30  
 : Sequence 30, Application US/29177155A  
 : Patent No. 6426205  
 : GENERAL INFORMATION:  
 : APPLICANT: Tyers, Mike  
 : APPLICANT: Williams, Andrew  
 : TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING BRICUTIN  
 : FILE REFERENCE: 11757.10USU1  
 : CURRENT APPLICATION NUMBER: US/09/177.165A  
 : CURRENT FILING DATE: 1998-10-22  
 : PRIOR APPLICATION NUMBER: 60/092.443  
 : PRIOR FILING DATE: 1998-07-10  
 : PRIOR APPLICATION NUMBER: 60/045.254  
 : PRIOR FILING DATE: 1997-10-24  
 : NUMBER OF SEQ ID NOS: 50  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 30  
 : LENGTH: 640  
 : TYPE: PRT  
 : ORGANISM: Saccharomyces cerevisiae  
 : US-09-177-165A-30

Query Match 4.7% Score 82; DB 4; Length 640;  
 Best Local Similarity 21.1%; Pred. No. 3;  
 Matches 39; Conservative 23; Mismatches 65; Indels 58; Gaps 7;  
 QY 62 NPQGNPKVNGAEGDVLAVYIESMLPRGVDPYGTICAMIRIFUGLGTGDATMCDPLPE 141  
 DB 450 SPQGDPTMTDCADESDTPSNEQEVLD-----NTPYPT 523  
 QY 142 KVRMIKLDSEKVKYKSRHTLPYKPHLGTLSVSEIUSNSLTPONHGN-----HGGNMDVP 194

DB 524 HLLSGGLNTIKLVNKTGKIRIQFG-----EVEGVWDIAADNFRILSGSHSSIKKW 577  
 QY 195 DISPGSITYPLVRAPGSRIL-----FQGD-----AHACGGEGHICGTAVEPASITT 239  
 DB 578 DLOGSKMHTF---NORRLQRETQHTQSLGDKVAPIACVCGIDSE-CFSGDERGCVAK 633  
 QY 240 IKVDL 244  
 DB 534 YKFDL 638

RESULT 11  
 US-08-476-062A-42  
 : Sequence 42, Application US/08476062A  
 : Patent No. 5872755  
 : GENERAL INFORMATION:  
 : APPLICANT: Arnaud, M. Amir  
 : TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY  
 : RESPONSES WITH BETA2 INTEGRINS  
 : NUMBER OF SEQUENCES: 53  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Fish & Richardson P.C.  
 : STREET: 225 Franklin Street  
 : CITY: Boston  
 : STATE: MA  
 : COUNTRY: US  
 : ZIP: 02110-2804  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette  
 : COMPUTER: IBM Compatible  
 : OPERATING SYSTEM: Windows95  
 : SOFTWARE: FastSeq for Windows Version 2.0  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/476,062A  
 : FILING DATE: 07-JUN-1995  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 09/216,081  
 : FILING DATE: 21-MAR-1994  
 : APPLICATION NUMBER: 07/637,830  
 : FILING DATE: 04-JAN-1991  
 : APPLICATION NUMBER: 07/535,842  
 : FILING DATE: 18-JUN-1990  
 : APPLICATION NUMBER: 07/212,573  
 : FILING DATE: 28-JUN-1988  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Freeman, John W.  
 : REGISTRATION NUMBER: 29,066  
 : REFERENCE/DOCKET NUMBER: 00786/066003  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 617/542-5070  
 : TELEFAX: 617/542-8906  
 : TELEX: 200154  
 : INFORMATION FOR SEQ ID NO: 42:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1170 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : FRAGMENT TYPE: internal  
 : US-08-476-062A-42

Query Match 4.7% Score 81.5; DB 2; Length 1170;  
 Best Local Similarity 22.7%; Pred. No. 9.4;  
 Matches 70; Conservative 35; Mismatches 109; Indels 95; Gaps 18;  
 QY 14 VGARKKPVTHLTIEEMOKERH-YTIGPYSTPVLTIIEPGRIIVDTDAFEAGAINSEQDIP 72  
 DB 284 IGIGK---HFQTKESQETLHKFKASKPASE-----FKILDIIF----- 318  
 QY 73 SOLLKMPFLNPNQNGPINKVNGAEGDVLAVYIE-----SNLPRGVDPYGTICAMIRIFUG 125  
 DB 319 --KUKDLFIERQKKIYVIEGTISKQCLTSFNMEI.SSSGISADLSRGNVAVGAKKAWGG 376

QY 126 LTGDTLTMG-----NDPLPEKVRMIKLDSEKVVW--SKRST-----LPYKPHITL 170  
 DB 377 F--LDLKAQLQDDTFIGNETPTPEVRAGYL-GYTVTWLPSRQKTSLLASGAPFYOHMGKV 433  
 QY 171 SVSPE-----IUSNSLTPONH-GGNKDVDPDG-POSITYPLVRAP-----GSR 212  
 DB 434 LLFOEPGGGCHWSQVOTHTGTQGSYFGGELCGVDVQDGTETELLIGAPFYGQRGR 493  
 QY 213 LFI-----GDAHACQSD-GEICGTAVERFASITTIKVDLIKNNKLSWPRMENAEKMS 263  
 DB 494 VFIYORRQIGFEFVSELQDPPGVFLG---RFGESAITALTJ-----NEKGLVD 538  
 QY 264 IGSARPLED 272  
 DB 539 VAVGAPLEE 547

## RESULT 12

PCT-US96-01114-42

Sequence 42, Application PC/TUS9601314  
 GENERAL INFORMATION:  
 APPLICANT: M. Amin Arnaut  
 TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN  
 TITLE OF INVENTION: ANTAGONISTS  
 NUMBER OF SEQUENCES: 78  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 COMPUTER: IBM PS/2 Model 502 or 55SX  
 OPERATING SYSTEM: MS-DOS (Version 5.0)  
 SOFTWARE: WordPerfect (Version 5.1)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US96/01314  
 FILING DATE: 30-JAN-96  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/380,167  
 FILING DATE: 30-JAN-95  
 ATTORNEY/AGENT INFORMATION:  
 NAME: John W. Freeman  
 REGISTRATION NUMBER: 29,066  
 REFERENCE/DOCKET NUMBER: G0786/267001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 542-5070  
 TELEFAX: (617) 542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 42:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1170  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear

PCT-US96-01114-42

Query Match 4.7% Score 82.5; DB 5; Length 1170;  
 Best Local Similarity 22.7%; Pred. NC. 9.4;  
 Matches 70; Conservative 35; Mismatches 105; Indels 95; Gaps 18;

QY 14 VAGRKPVTHHTEEMQKEPH-VTCIPYSTIVLIEPGRKLIIVTRAFEGAINSEODIP 72  
 DB 284 IGIGK----HFQTKESQETJHFKASKPASE-----EVKILDTF----- 316  
 QY 73 SOLIKMPFLPNQPIWVNGAEKGEVLAVYIP-----SMIPKSVDPYGTICAMIHFGS 127  
 DB 319 --KLKDLFTERQKKIVLECTSKQDLTSFNMELSSSGISACLRSCHAVAVGAKWAG 375  
 QY 126 LTGDTLTMG-----NDPLPEKVRMIKLDSEKVVW--SKRST-----LPYKPHITL 170

DB 377 F--LDLKAQLQDDTFIGNETPTPEVRAGYL-GYTVTWLPSRQKTSLLASGAPFYOHMGKV 433  
 QY 171 SVSPE-----IUSNSLTPONH-GGNKDVDPDG-POSITYPLVRAP-----GSR 212  
 DB 434 LLFOEPGGGCHWSQVOTHTGTQGSYFGGELCGVDVQDGTETELLIGAPFYGQRGR 493  
 QY 213 LFI-----GDAHACQSD-GEICGTAVERFASITTIKVDLIKNNKLSWPRMENAEKMS 263  
 DB 494 VFIYORRQIGFEFVSELQDPPGVFLG---RFGESAITALTJ-----NEKGLVD 538  
 QY 264 IGSARPLED 272  
 DB 539 VAVGAPLEE 547

RESULT 13  
 US-07-978-895-4  
 Sequence 4, Application US/0797895  
 Patent No. 5480968  
 GENERAL INFORMATION:  
 APPLICANT: Kraus, Mathias H.  
 TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE  
 TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND  
 TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THERETO  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sullie 400  
 STREET: 133 Carnegie Way, N.W.  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: U.S.A.  
 ZIP: 30303  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/978,895  
 FILING DATE: 19921110  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/444,406  
 FILING DATE: 01-DEC-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Perryman, David S.  
 REGISTRATION NUMBER: 33,435  
 REFERENCE/DOCKET NUMBER: 1414-029  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (404) 688-0770  
 TELEFAX: (404) 688-9660  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1342 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-978-895-4

Query Match 4.6% Score 81; DB 1; Length 1342;  
 Best Local Similarity 24.6%; Pred. No. 13;  
 Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY 5 IMAKRGVAGRKPVTHHTEEMQKEPHYITIGPYSITPVLIEPGRKLIIVTRAFEGAIN 67  
 DB 978 KRESGPGIANGSPHETLNKK:EEVE--DEPEDLDLDUEAEEDNAT-----TDS 1029  
 QY 68 KQDIPSOLEKMP-----FLAFONGPIWVNGAEKGEVLAVYIESKL-----PKGVY 112  
 DB 1030 ALSLVIGTINPSSGSLSPSSGYMPMGNLGCSCG---ESAVGSSSESGCPWPVSLRP 1096  
 QY 113 -PYGICAMIPFGSLTGIDLTAKLNDPLPEKVRMIKLDSEKVVW--SKRST-----EVKILDTF 161

```

Db      1087 MPRCLASSESEGVGTGSEA-----ELQKVMCRSRSSRSPRPGDSAYHSORHSL 1140
QY      162 -PYKPHIGTILSVSP---EIDINS-LTPDNH 187
Db      1141 TPVTP-----LSPGLEEDVNGVYVMDTH 1165

RESULT 14
US-08-473-119-4
; Sequence 4, Application US/08473119
; Patent No. 5820959
; GENERAL INFORMATION:
; APPLICANT: Kraus, Matthias H.
; APPLICANT: Aaronson, Stuart A.
; TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND
; TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THERETO
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 400
; STREET: 133 Carnegie Way, N.W.
; City: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/473.119
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/978,895
; FILING DATE: 10-NOV-1992
; APPLICATION NUMBER: US 07/444,406
; FILING DATE: 07-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-473-119-4

Query Match: 4.6%; Score 81; DB 2; Length 1342;
Best Local Similarity 24.6%; Pred. No. 13;
Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY      8 IMAKRGVAGKPKVTHLIEEMCKEFTYIGYSTPVLTIETPGDRIIVDTCAFEKAINS 67
Db      978 IKRESGPGIAPGPEPHGLTNKKLEVE--LEPELDLDLEAEEDNLAT-----TLGS 1029
QY      68 EODIPSQLKMP-----FLNPQNGPIVNGAEKGVLAIVIESML-----PRGVD--- 112
Db      1030 ALSPLVGTNLNPRGSQLSLSPSSGYMPMNOGLGESCQ---ESAVSGSSERCPRPVSLHP 1086
QY      113 -PYGICAMIPHFGGLTGTDLTAMLDNLPPEKVRMKLDS-----EKVYWSKRHTL- 161
Db      1087 MPRCLASSESEGHVTGSEA-----ELQKVMCRSRSSRSPRPGDSAYHSORHSL 1140
QY      162 -PYKPHIGTILSVSP---EIDINS-LTPDNH 187
Db      1141 TPVTP-----LSPGLEEDVNGVYVMDTH 1165

```

Search completed: July 3, 2003, 15:50:45  
Job time : 38 secs

```

Db      1141 TPVTP-----LSPGLEEDVNGVYVMDTH 1165

RESULT 15
US-08-475-352-4
; Sequence 4, Application US/08475352
; Patent No. 5916755
; GENERAL INFORMATION:
; APPLICANT: Kraus, Matthias H.
; APPLICANT: Aaronson, Stuart A.
; TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND
; TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THERETO
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 400
; STREET: 133 Carnegie Way, N.W.
; City: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/475.352
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/978,895
; FILING DATE:
; APPLICATION NUMBER: US 07/444,406
; FILING DATE: 07-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-475-352-4

Query Match: 4.6%; Score 81; DB 2; Length 1342;
Best Local Similarity 24.6%; Pred. No. 13;
Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY      8 IMAKRGVAGKPKVTHLIEEMCKEFTYIGYSTPVLTIETPGDRIIVDTCAFEKAINS 67
Db      978 IKRESGPGIAPGPEPHGLTNKKLEVE--LEPELDLDLEAEEDNLAT-----TLGS 1029
QY      68 EODIPSQLKMP-----FLNPQNGPIVNGAEKGVLAIVIESML-----PRGVD--- 112
Db      1030 ALSPLVGTNLNPRGSQLSLSPSSGYMPMNOGLGESCQ---ESAVSGSSERCPRPVSLHP 1086
QY      113 -PYGICAMIPHFGGLTGTDLTAMLDNLPPEKVRMKLDS-----EKVYWSKRHTL- 161
Db      1087 MPRCLASSESEGHVTGSEA-----ELQKVMCRSRSSRSPRPGDSAYHSORHSL 1140
QY      162 -PYKPHIGTILSVSP---EIDINS-LTPDNH 187
Db      1141 TPVTP-----LSPGLEEDVNGVYVMDTH 1165

```



\_\_\_\_\_

SeqCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM nucleotide - nucleic search, using sw mode!

Run on: July 6, 2003, 04:04:05 : Search time 1972 seconds  
(without alignments)  
11842.744 Million cell: updates/sec

Title: US-10-086-082-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgtggcc.....aatgcatttcatttgatccc 1442

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 42302132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:

1: em\_estba:\*\*\*  
2: em\_estnum:\*\*\*  
3: em\_estin:\*\*\*  
4: em\_estov:\*\*\*  
5: em\_estov:\*\*\*  
6: em\_estov:\*\*\*  
7: em\_estov:\*\*\*  
8: em\_hic:\*\*\*  
9: gb\_est1:\*\*\*  
10: gb\_est2:\*\*\*  
11: gb\_hic:\*\*\*  
12: gb\_est3:\*\*\*  
13: gb\_est4:\*\*\*  
14: gb\_est5:\*\*\*  
15: em\_estfun:\*\*\*  
16: em\_estem:\*\*\*  
17: gb\_gss:\*\*\*  
18: em\_gss:\*\*\*  
19: em\_gss\_inv:\*\*\*  
20: em\_gss\_pln:\*\*\*  
21: em\_gss\_vrt:\*\*\*  
22: em\_gss\_fur:\*\*\*  
23: em\_gss\_mdm:\*\*\*  
24: em\_gss\_mas:\*\*\*  
25: em\_gss\_othe:\*\*\*  
26: em\_gss\_pro:\*\*\*  
27: em\_gss\_rod:\*\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.6	2.7	280	10 AW380638	AW380638 RC2-HT027
2	37.8	2.6	551	13 BM615374	BM615374 170006871
3	36.2	2.5	533	10 AV551507	AV551507 AV551507
4	36.2	2.5	678	13 BM582997	BM582997 170006872
5	36.2	2.5	1101	17 CNS00370	ALC64465 Drosophila
6	36	2.5	495	10 AV629284	AV629284 AV629284

C	7	35.6	2.5	482	17	BB767573	BB767573 BMRAC348F
C	8	35.6	2.5	735	17	AG117547	AG117547 Pap. 1509
C	9	35.2	2.4	612	10	BE442656	BE442656 WHELI111P
C	11	35.2	2.4	772	17	BB709682	BB709682 BOWE1171F
C	12	35	2.4	268	9	AA020713	AA020713 ze63L10.8
C	13	35	2.4	421	14	H75593	H75593 yuc5104.1
C	14	35	2.4	441	9	AA019639	AA019639 ze6241.8
C	15	35	2.4	495	9	AA061557	AA061557 ze45d06.8
C	16	35	2.4	505	12	BG834211	BG834211 352349 MA
C	17	34.8	2.4	631	17	B57961	B57961 CIT-HSP-201
C	18	34.8	2.4	328	9	AG109318	AG109318 AJ193318
C	19	34.8	2.4	396	17	AG010392	AG010392 HS_5101-B
C	20	34.8	2.4	422	17	AG055596	AG055596 HS_5229-A
C	21	34.8	2.4	475	9	AI743670	AI743670 w41e06.X
C	22	34.8	2.4	477	9	AI346857	AI346857 qp53e12.X
C	23	34.8	2.4	482	14	BQ247372	BQ247372 Tab250473
C	24	34.8	2.4	502	9	AI289776	AI289776 qw12a01.X
C	25	34.8	2.4	524	12	BG052847	BG052847 KH122.14
C	26	34.8	2.4	511	17	AG048299	AG048299 RPT1-11-2
C	27	34.4	2.4	562	17	AG047896	AG047896 RPT1-11-2
C	28	34.4	2.4	474	10	BE498530	BE498530 WHE0971-C
C	29	34.4	2.4	483	9	AG033827	AG033827 os62c06.8
C	30	34.4	2.4	491	14	H86687	H86687 y193a12.r1
C	31	34.4	2.4	623	17	B59197	B59197 CIT-HSP-205
C	32	34.4	2.4	641	13	BM392441	BM392441 50071-2-1
C	33	34.4	2.4	641	13	BM393654	BM393654 50072-2-1
C	34	34.4	2.4	100	17	CNS00020	AL077401 Drosophila
C	35	34.2	2.4	466	13	BM136789	BM136789 WHE2617-E
C	36	34.2	2.4	531	13	B1246954	B1246954 B7246954
C	37	34.2	2.4	656	13	B2234037	B2234037 B2234037
C	38	34.2	2.4	661	10	BE489900	BE489900 WHE3363-H
C	39	34.2	2.4	503	13	EM018122	EM018122 603645981
C	40	34	2.4	347	14	R03461	R03461 y108a12.s1
C	41	34	2.4	515	12	BE252201	BE252201 EST419463
C	42	34	2.4	540	12	BE252018	BE252018 EST419290
C	43	34	2.4	565	17	AG438850	AG438850 HS_5103-A
C	44	34	2.4	569	17	BE251067	BE251067 EST418324
C	45	34	2.4	760	12	B5574169	B5574169 602595048

#### ALIGNMENTS

RESULT 1  
AW380638  
LOCUS RC2-HT0275-01-199-011-111 HT0275 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION AW380638  
VERSION AW380638.1 GI:6885297  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Pekaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 280)  
HCCP <http://www.ludwig.org.br/ORESTES>.  
The PAPESP/LICR Human Cancer Genome Project  
Unpublished (1999)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: -55-11-2704922  
Fax: -55-11-2707000  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC2&t2=RC2-HT0275-01199-011-flist&t3=1999-11-01&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 37







1-7-22 Suehiro-chou, Tsukuba-shi, Ibaraki, Japan  
(E-mail: chimbas@cc.riken.go.jp, URL: http://hjb-acc.riken.go.jp/)  
Tel: 81-45-503-9111, Fax: 81-45-503-9170  
Clones are derived from the chimpanzee BAC library FIB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

PRIMERS  
Sequencing: -21M13

LIBRARY  
Vector : PKS145  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. 735  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-125G18.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
188 a 186 c 160 g 198 t 3 others

BASE COUNT  
ORIGIN  
Query Match 2.58; Score 35.6; DB 17; Length 755;  
Best Local Similarity 58.58; Pred. No. 13;  
Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
QY 1315 TACTCTTCAGGAGGAGTGGCGGCGGAGAGTGTGCACAAATGGCGGGGAGGACCCAG 1374  
DB 123 TGCCTTCAGGCGGAGGAGTGGCGGCGGAGAGTGTGCACAAATGGCGGGGAGGACCCAG 1374  
QY 1375 GCTATTCGCGAAATTAATCAATATGGCGGAGGAGTGTGCACAAATGGCGGGGAGGACCCAG 1422  
DB 163 GTTGTGACCAATTAATTAATGGCGGAGGAGTGTGCACAAATGGCGGGGAGGACCCAG 228

RESULT 9  
BE442656  
LOCUS BE442656 611 bp mRNA linear EST 25 JUL-2000  
DEFINITION Wheat etiolated seedling root normalized cDNA  
library Triticum aestivum cDNA clone WHE1101.D03.G05, mRNA  
sequence.  
ACCESSION BE442656  
VERSION BE442656.1 GI:9442089  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae  
; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 611)  
AUTHORS Anderson, O.D., Chao, S., Choi, E.W., Close, T.J., Fiedler, R.D., Han  
P.S., Hsu, C.C., Kang, Y., Lazo, S.K., Miller, S., Nguyen, H.T.,  
Rausch, C., Seaton, C., Tong, J.C. and Zhang, B.  
TITLE The structure and function of the expressed portion of the wheat  
genomes - Normalized root cDNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: andersn@nps.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stragene SK primer.  
Location/Qualifiers  
1. 611  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4563"

FEATURES  
source  
1. 611  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4563"

/clone="WHE1101.D03.G05"  
/clone\_lib="Wheat etiolated seedling root normalized cDNA  
library"  
/tissue\_type="Root"  
/dev\_stage="Five day old etiolated seedling"  
/lab\_host="E. coli DH10B"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid  
pBluescript SK; Site 1: EcoRI; Site 2: XhoI; Seeds were  
surface-sterilized, germinated and grown aseptically in  
the dark at room temperature on filter paper with water,  
nystatin and celastrol in covered crystallization  
dishes. Roots were harvested. The tissue, total RNA, and  
poly(A) RNA were prepared, a cDNA library was made in the  
TJ Close lab (Choi, Close, Fenton) at the University of  
California, Riverside. The cDNA clones were in vivo  
excised to give pBluescript phagemids before  
normalization was carried out. The mass excision of  
phagemid library and normalization were done in HT Nguyen  
lab by D. Zhang at Texas Tech University. Normalization  
protocol used was that of Soares. Plasmid DNA  
preparations and DNA sequencing were performed in the OD  
Anderson lab (all other authors)."  
BASE COUNT 181 a 111 c 153 g 166 t  
ORIGIN

Query Match 2.48; Score 35.2; DB 10; Length 611;  
Best Local Similarity 51.28; Pred. No. 15;  
Matches 82; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
QY 719 GAATTTGACTCAATCAATTCAGTACGCCAGACATCAACGGGGGAGTATGGATGGCG 778  
DB 141 GCNATGAGCAGGAGGAGTATTCCTTGGAGAGAGCAATGGGGGAGTATTCGACATTAAG 200  
QY 779 GATATGAGCAGGAGGAGTATTCCTTGGAGAGAGCAATGGGGGAGTATTCGACATTAAG 848  
DB 231 AATCTAGCAGGAGGAGTATTCCTTGGAGAGAGCAATGGGGGAGTATTCGACATTAAG 260  
QY 839 ATTGTGATGCCAGTATTCCTTGGAGAGAGCAATGGGGGAGTATTCGACATTAAG 878  
DB 241 ACTGCGATATGCACTTTCCAGGAGGAGTATTCGACATTAAG 300

RESULT 10  
BE442656  
LOCUS BE442656 772 bp cDNA linear GSS 20-FEB-2002  
DEFINITION BE442656.1 Brassica oleracea genomic clone BOMEL07F, cDNA  
sequence.  
ACCESSION BE442656  
VERSION BE442656.1 GI:977569  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 772)  
AUTHORS Town, C.D., Van Aken, S., Lutterback, T. and Fraser, C.M.  
TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOMEL07F  
Contact: Chris Town  
77GR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@stg.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: 1P  
Class: sheared ends.  
Location/Qualifiers  
1. 772  
/organism="Brassica oleracea"  
/strain="Tol000DH3"

FEATURES  
source  
1. 772  
/organism="Brassica oleracea"  
/strain="Tol000DH3"

FEATURES  
source  
1. 772  
/organism="Brassica oleracea"  
/strain="Tol000DH3"

FEATURES  
source  
1. 772  
/organism="Brassica oleracea"  
/strain="Tol000DH3"

FEATURES  
source  
1. 772  
/organism="Brassica oleracea"  
/strain="Tol000DH3"

```

/db_xref="taxon:3712"
/clone="BOMEL07"
/clone_lib="BC2_3_KB"
/note="Vector: pHOS1; Site1: 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      175 a 165 c 181 g 251 t
ORIGIN
Query Match      2.4% Score 35.2; DB 17; Length 772;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 592 GGCGATGTCATATCCGTCGCCAGAAAGCTGCCGATGATTAACCTCGACACGAAAA 651
DB 636 GACATCCCGATAGACGGTCCATATACACGAGATTCCTTCTAGAAATCATAAAA 577
QY 652 GGCTACTGGAGCAACGCTATACCTCCCTATTAACCCATATTCGACCTTGAGCGT 711
DB 576 COTCTAAGTCGAAAGGCGCAAGAGAGCGCTAGAAAGCGGCTAACAGCCACATATAT 517
QY 712 ATCCGCGAGAAATGACATCAATCAATTCATCAGCGCGAGACATACGCGCGGAGTA 767
DB 516 TTATACGAAATGAGCCATCATGCTGATGACGCTTATCTTTACTTCGCGAAAA 461

RESULT 11
AAQ20733 268 bp mRNA linear EST 30-JAN-1997
LOCUS z663b10.s1 Soares retina N264Rk Homo sapiens cDNA clone
DEFINITION IMAGE:363643 3' similar to WP:K02423 C002791 BOMBATINDE-SENSITIVE
NA-K-Cl cotransporter 1, mRNA sequence.
AAQ20733
ACCESSION AAQ20733.1 GI:1484516
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 268)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapella,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisan,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-826 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63109
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free thro. g. LINCS: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2350 Std Error: 0.00
Seq primer: -40M13 fwd. from Areshan
High quality sequence stop: 171.
Location/Qualifiers
1..268
/organism="Homo sapiens"
/db_xref="Gene:1280346"
/db_xref="taxon:9606"
/clone="IMAGE:363643"
/clone_lib="Soares retina N264Rk"
/sex="male"
/tissue="retina"
/der_stage="55 year old"
/lab_host="JELOR (ampicillin resistant)"
/note="Organ: eye; Vector: pMT7.0 (Pharmacia) with a
modified polylinker; Site1: Not 1; Site2: Eco RI; 1st
strand cDNA was primed with a Not 1 - oligo(dT) primer (5'
TCTTACCAATCTGAATGGAGCGCGCGCGCTTTTITTTTITTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pMT7.0 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)-RNA was extracted 5
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Benito
Soares and M.Fatima Bonaldo."
BASE COUNT      50 a 72 c 71 g 71 t
ORIGIN
Query Match      2.4% Score 35; DB 9; Length 268;
Best Local Similarity 50.6%; Pred. No. 9.8;
Matches 43; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 535 CTACGGCATCTGGCGCATGATCCGCAATTTGGCGACATGACGGAGCGACCTGACGGC 594
DB 167 CAAGGACTCTTACAGCTGAGCCCGGCTGGGAGAACTTGAACCACTCCAAAGTGGGCG 108
QY 595 CATGCTCAATGATCGCTGCGAGAAAGGTGCGCATGATTAACCTGACATGAGAAAGT 654
DB 107 CATGCACACTGCTGCGCTGAACGAGGTGCTGTAAGAAATCCCGGAGCGGCAAGCT 48
QY 655 CTACTGGAGCAAGCGCATACGCTTCCTTATTAACGCCATATG 693
DB 47 TGTTTTCCTACATGCTTGGGCTGCTGCGCAACGCAATGGTG 4

RESULT 12
H75593/c 421 bp mRNA linear EST 01-NOV-1995
LOCUS J005104.r1 Soares fetal liver spleen cDNA clone
DEFINITION IMAGE:232927 5' mRNA sequence.
H75593
ACCESSION H75593.1 GI:1049521
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 421)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapella,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisan,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-826 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63109
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert size: 1012
High quality sequence stop: 366
Source: IMAGE Consortium, JNL
This clone is available royalty-free through LINCS: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1012 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 366.
Location/Qualifiers
1..421
/organism="Homo sapiens"
/db_xref="Gene:3786104"
/db_xref="taxon:9606"
/clone="IMAGE:232927"
/clone_lib="Soares fetal liver spleen cDNA"
FEATURES
source

```

```
/sex="male"
/dev stage="23 week post conception fetus"
/lab host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pTZ19 (Pharmacia)
with a modified polylinker; Site: 1; Pac 1; Site 2: Eco RI;
1st strand cDNA was primed with a Pac 1 - oligo(3T) primer [5'
5' AACTGGAGAATTAATAAGATCTTTTITTTTITTTT 3'].
double-stranded cDNA was ligated to Eco RI adapters
(Pharmacia), digested with Pac 1 and cloned into the Pac 1
and Eco RI sites of the modified pTZ19 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."
```

```
/sex="male"
/tissue type="retina"
/dev stage="55 year old"
/lab host="DH10B (ampicillin resistant)"
/notes="Organ: eye; Vector: pTZ19 (Pharmacia) with a
modified polylinker; Site: 1; Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not 1 - oligo(GT) primer [5'
5' GTTACCAATCTGAAGTGGAGGCGCGCTTTTITTTTITTTT 3'].
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not 1 and cloned into
the Not 1 and Eco RI sites of a modified pTZ19 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)- RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M. Fatima Bonaldo."
```

```
BASE COUNT 89 a 104 c 107 g 113 t 5 others
ORIGIN
Query Match 2.4% Score 35; DB 14; Length 421;
Best Local Similarity 51.6% Pred. No. 13;
Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 1044 ACTGGCTGGTAGAAGACTTTGGCTTCGACAAATGGGATGGCTACATCTCTGAGTCAAT 1103
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 242 ACTGGGAAGAGATGACTTTGGCTGCGAGGACATGATCTCAACGACGCTCTGAGACAT 183
QY 1104 CGGCAAGTCGCGCTGGGCAACATGCTGACCCCAATACACGCTTGGCGGATCTGA 1163
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 182 CCCACAGGAGTCTGCTGGGCGGATAGCCGTGACAGTTACACCGCTTGGGGTGTAAC 123
QY 1164 ACAAAACCTGTTAGTTAGTAGGAATAAATAAC 1198
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 122 ACTATACCTTATATAGACGCTGATCTTTCTACCC 88
```

```
Query Match 2.4% Score 35; DB 14; Length 421;
Best Local Similarity 51.6% Pred. No. 13;
Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
```

```
QY 1044 ACTGGCTGGTAGAAGACTTTGGCTTCGACAAATGGGATGGCTACATCTCTGAGTCAAT 1103
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 242 ACTGGGAAGAGATGACTTTGGCTGCGAGGACATGATCTCAACGACGCTCTGAGACAT 183
QY 1104 CGGCAAGTCGCGCTGGGCAACATGCTGACCCCAATACACGCTTGGCGGATCTGA 1163
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 182 CCCACAGGAGTCTGCTGGGCGGATAGCCGTGACAGTTACACCGCTTGGGGTGTAAC 123
QY 1164 ACAAAACCTGTTAGTTAGTAGGAATAAATAAC 1198
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 122 ACTATACCTTATATAGACGCTGATCTTTCTACCC 88
```

```
RESULT 14
AA019639/c 441 bp mRNA linear EST 10-JAN-1997
LOCUS ze46d06.s1 Soares retina N2B4HR Homo sapiens cDNA clone
DEFINITION IMAGE:362027 3' similar to WP:K02A2.3 CE02791 BUNSIANTIDE-SENSITIVE
NA-K-CL cotransporter ; mRNA sequence.
```

```
AA019639 GI:1483067
VERSION 1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE 1 (bases 1 to 441)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
```

```
1 (bases 1 to 441)
Hillier, L., Lennarz, S., Beckert, M., Bonaldo, M.F., Chiswick, H.,
Chiswick, S., Dietrich, K., DuBoque, T., Pavello, A., Gish, W., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Margis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thirumangalakudi, V., Trevisan, R.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, R.,
Generation and analysis of 260,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
57044479
```

```
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
```

```
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2010 Std Error: 0.00
Seq primer: 40M13 fwd. from Amersham
High quality sequence stop: 360.
Location/Qualifiers
1..441
```

```
FEATURES
Source
/organism="Homo sapiens"
/db_xref="CDB:1280303"
/db_xref="EAX:9606"
/clone="IMAGE:362027"
/clone_lib="Soares retina N2B4HR"
```

```
BASE COUNT 65 a 137 c 120 g 113 t 5 others
ORIGIN
Query Match 2.4% Score 35; DB 9; Length 441;
Best Local Similarity 50.6% Pred. No. 14;
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 535 CTACGGATCTGCCCATGATTCGCGCAATTTTGGCGGACTGACCGGGACCGACCTGACGGC 594
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 167 CAAGGACTCTTTCATCATGAAGCGGAGTGGGAGAACTTGAACCACTCCCACTGGCGG 128
QY 595 CATGCTCAATGATGCGTGGCGGAGAAAGTGGCGCATGATTAAAGTCGACAGTGAAGAAGT 654
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 107 CATGCACTGCTGCGTGGCGGAGTGGCGCATGATTAAAGTCGACAGTGAAGAAGT 48
QY 655 CTACTGAGCAAGCAAGCAAGCTTCCTATTAACCCCATATTG 698
||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| ||||| 1 ||||| |||||
Db 47 TGTITTCCTACATGCTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGT 4
```

```
AA01557 492 bp mRNA linear EST 29-NOV-1996
ze46d06.s1 Soares retina N2B4HR Homo sapiens cDNA clone
IMAGE:362027 3' similar to WP:K02A2.3 CE02791 BUNSIANTIDE-SENSITIVE
NA-K-CL cotransporter ; contains Alu repetitive element.; mRNA
sequence.
```

```
AA01557 GI:1437022
VERSION 1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE 1 (bases 1 to 492)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
```

```
1 (bases 1 to 492)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Lacy, M., Lennarz, S., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, R.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, R.,
Generation and analysis of 260,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
57044479
```

```
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
```

```
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 3433 Std Error: 0.00
Seq primer: mob.REGA-ET
High quality sequence stop: 233.
Location/Qualifiers
1..492
```

```
FEATURES
Source
/organism="Homo sapiens"
/db_xref="GDB:1278730"
```



```

/db_xref="taxon:9606"
/clone="IMAGE:362027"
/clone_lib="Soares retina N2b43R"
/sex="male"
/tissue_type="retina"
/dov_stage="55 year old"
/lab_host="JH10B (ampicillin resistant)"
/notes="Organ: eye; Vector: p1730 (Pharmacia); with a
modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st
strand cDNA was primed with a Not 1 - oligo(dT) primer; 5'
TGTTACCACTGTGAGTGGAGCGCGCGCTTTCTTTTITTTI 3'), RI
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not 1 and cloned into
the Not 1 and Eco RI sites of a modified pT73 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)- RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D., Ph.D., from the
University of Toronto. Library constructed by Sento
Soares and M.Fatima Bonaldi."
BASE COUNT      80 a  141 c  131 g  125 t      11 others
ORIGIN
Query Match      2.4%; Score 35; DB 9; Length 492;
Best Local Similarity 50.6%; Pred. No. 15;
Matches 83; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 535 CTACGGGATCTCGGCATATTCGCCATTTTGGGCGACTGACGGGAGTGGAGGAGG 594
DB 167 CAAGGACATCTTCAGCATGAACCGGAGTGGAGAACTGAACGAGTCCACAGTCCGCG 108
QY 595 CATGCGCAATGATCGCTGCCAGAAAGTGGCGGATGATTAAATCGACAGTGAAGG 654
DB 107 CATGCACATGCTGCGGCTGATGACGAGGATCGTGAAGAAATCCCGGAGCGAAAGCT 48
QY 555 CTATGGAGCAACGCCAATCCCTCCCTATTAACCTCATTTG 598
DB 47 TGTTCGCTCAACATGCTCGGCGCTCCCGCAACCGCAATG 4

RESULT 15
RG834211
LACUS      505 bp      mRNA      150bp      EST 25 MAY-2001
DEFINITION 352349 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  RG834211
VERSION     RG834211.1 GI:14199291
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 505)
AUTHORS   Fahrenkrug, S.C., Freking, B.A., Kohrer, G.A., Smith, J.P.L., Casas, D.,
            Stone, R.T., Heaton, M.P., Grosse, W.X., Bennett, G.A., Luedero, W.W.,
            and Keele, J.W.
            Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
            Unpublished (2000)
JOURNAL
COMMENT    Contract: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@mail.marc.usda.gov
            Single pass sequencing. Bases called and aligned with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAACAGCTATGACCAT
            BACKWARD: GTTTCGCCAGTCCAGAG
            Plate: 113 row: 1 column: 2
            Seq primer: ATTAGTGCAGCTATAG.
            Location/Qualifiers

```

```

Source
1..505
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="JH10B"
/notes="Vector: pCMV SPORT5; Site_1: Not 1; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT      104 a  121 c  158 g  122 t
ORIGIN
Query Match      2.4%; Score 35; DB 12; Length 505;
Best Local Similarity 55.3%; Pred. No. 15;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 2 CGGGGAATCCCATGTGGCGGTGATCTGTCGAGGAGGATATTGGATGATCCAGGGGC 61
DB 334 CTTCAATCTTCACTGCTGCTCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 393
QY 62 CGCACAGCGCTGTGGGTAATGATTAAGGCGCTGTTTGTAGAAAGCTGACCCACACA 121
DB 194 TGAATACGTGTGAGGACCTGACCGTGGCCCTGCGAGGAGGAGGAGCTGACCTCCACA 453
QY 122 GCT 124
DB 454 GCT 456

Search completed: July 6, 2003, 06:40:28
Job time : 1978 secs

```

GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: July 6, 2003, 04:47:45 : Search time 97 seconds  
 (without alignments)  
 4559.050 Million cell updates/sec

Title: US-10-056-062-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgtagcc.....aatgcaatttcattgacac 1442

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum US seq length: 0

Maximum DB seq length: 2000000200

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37.5	2.6	503	1	US-08-123-343A-15
C 2	37.5	2.6	1233	4	US-09-431-573-1
C 3	37.6	2.6	1233	4	US-09-431-573-2
C 4	37.6	2.6	1233	4	US-09-431-573-3
C 5	37.6	2.6	3900	1	US-08-123-343A-6
C 6	35.2	2.4	429	4	US-09-144-367-7
C 7	34.2	2.4	5733	2	US-08-473-533A-1
C 8	33.2	2.3	3201	2	US-08-533-770A-1
C 9	33.2	2.3	4726	1	US-08-533-770A-2
C 10	32.8	2.3	7218	1	US-08-232-453-14
C 11	32.4	2.2	3176	1	US-09-308-572-3
C 12	32.4	2.2	9551	1	US-08-056-200-23
C 13	32.4	2.2	9551	2	US-08-800-644-53
C 14	32.4	2.2	1226	1	US-08-374-666-4
C 15	32.2	2.2	1349	1	US-08-374-666-1
C 16	32.2	2.2	6254	4	US-08-921-2-9-126
C 17	31.8	2.2	1917	4	US-08-943-731-163
C 18	31.8	2.2	20084	4	US-08-943-731-5
C 19	31.6	2.2	1380	4	US-09-513-783A-169
C 20	31.2	2.2	2167	3	US-08-743-637B-16
C 21	31.2	2.2	2167	3	US-08-528-840B-16
C 22	31	2.1	4403765	4	US-09-103-840A-2
C 23	31	2.1	4411529	4	US-09-103-840A-1
C 24	30.8	2.1	1187	5	PCT-US95-13536-2
C 25	30.8	2.1	1188	5	PCT-US95-13536-1
C 26	30.6	2.1	1542	4	US-09-202-893B-3
C 27	30.6	2.1	4403765	4	US-09-103-840A-2

C 28	30.6	2.1	4411529	4	US-09-103-840A-1
C 29	30.4	2.1	1049	3	US-08-652-643-1
C 30	30.2	2.1	747	4	US-08-958-416-1016
C 31	30.2	2.1	10095	3	US-08-822-586-45
C 32	30	2.1	1614	4	US-09-522-217-7
C 33	30	2.1	2615	3	US-09-940-905-1
C 34	29.8	2.1	283	4	US-09-007-005-17
C 35	29.8	2.1	239	4	US-06-244-796-17
C 36	29.8	2.1	417	4	US-08-936-165A-17
C 37	29.4	2.0	1588	1	US-08-485-107-1
C 38	29.4	2.0	7952	4	US-08-961-527-112
C 39	29.2	2.0	1524	4	US-09-134-031C-2055
C 40	29.2	2.0	45899	1	US-08-471-119A-1
C 41	29	2.0	300	1	US-08-083-946-6
C 42	29	2.0	300	3	US-08-452-915-6
C 43	29	2.0	615	4	US-05-134-001C-138
C 44	29	2.0	850	1	US-07-611-528A-1
C 45	29	2.0	850	1	US-08-083-946-1

## ALIGNMENTS

RESULT 1  
 US-08-123-343A-15/c  
 : Sequence 15, Application: US/98123343A  
 : Patent No. 5593879  
 : GENERAL INFORMATION:  
 : APPLICANT: Steller, Hermann  
 : APPLICANT: Adams, John M.  
 : APPLICANT: Greiner, Megan E.  
 : APPLICANT: White, Kristin  
 : TITLE OF INVENTION: Cell Death Genes of Drosophila  
 : NUMBER OF INVENTIONS: Molanogaster and Vertebrate Analogs  
 : NUMBER OF SEQUENCES: 15  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSES: Hamilton, Brook, Smith & Reynolds, P.C.  
 : STREET: Two Militia Drive  
 : CITY: Lexington  
 : STATE: MA  
 : COUNTRY: US  
 : ZIP: 02173  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent in Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/123,343A  
 : FILING DATE: 17-SEP-1993  
 : CLASSIFICATION: 800  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 06/104,957  
 : FILING DATE: 15-JAN-1993  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Granchan, Patricia  
 : REGISTRATION NUMBER: 32,227  
 : REFERENCE/DOCKET NUMBER: MIT-5907A  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 617-651-6240  
 : TELEFAX: 6186-9340  
 : INFORMATION FOR SEQ ID NO: 15:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 503 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: unknown  
 : TOPOLOGY: unknown  
 : MOLECULE TYPE: DNA (genomic)  
 : US-08-123-343A-15

Query Match 2.6% Score 37.6; DB 1: Length 503;  
 Best Local Similarity 48.6%; Pred. No. 0.02; 109; Indels 0; Gaps 0;  
 Matches 103; Conservative 0; Mismatches 109;





TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5733 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:

NAME/KEY: CDS  
LOCATION: 2434..5037  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5117..5467  
US-08-473-553A-1

Query Match 2.4%; Score 34.2; DB 2; Length 5733;  
Best Local Similarity 50.3%; Pred. No. 1.2;  
Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1136 CCCAATACACCGTTCGGCGGATGCTGACACAAACCTCTTASTTATAGCAATAAGTA 1195  
DB 3709 CTCACGGGACGCTTCGGCGGGGCTTTCAATCTACCGTTTATACGATTCGAGTTC 3766  
QY 1196 ACCGGTGAACATACCCGATGTAGATCGGGTAAATGTGTAATCAACAATGCTAAT 1255  
DB 3769 ACTGATAATTCCTCTCCGTGACCTCGGTACGATGTCGAGGAGTTCGAGTTCAG 3828  
QY 1256 TTACACGCTAATACGCTGCATATAGGCGGACAGATACACCATCAT 1302  
DB 3829 ATTACCTCTCTAACACAGCTTTCGGCGGAGATTCACCTCGCAT 3878

## RESULT 8

US-08-633-770A-3/c  
Sequence 3; Application US/08633770A  
Patent No. 5908760

## GENERAL INFORMATION:

APPLICANT: Bojsen, Kirsten  
APPLICANT: Yu, Shukun  
APPLICANT: Kragh, Karsten  
APPLICANT: Christensen, Torve  
APPLICANT: Marcussen, Jan  
TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,770A  
FILING DATE: July 8, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP94/03398  
FILING DATE: OCT-15-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY006.001APC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:

## INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3201 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-633-770A-3

Query Match 2.3%; Score 33.2; DB 2; Length 3201;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 243 CCGGGGTAAACCGGTAAACGATCACCCTGACGGAAGAAATGCAAAAGACTTTCATACA 302  
DB 1228 CTGGGCTCACTTTGAACGCTCAACAGACCTGTATAAAGTGTGATCAATTTCCGGGACC 969  
QY 303 CCATGGCCCTTATTCACACCCGCTCTGACCATCGAACCCGCTGACCGGATATCTCG 362  
DB 968 GTATCCGACGTGATACCGTAAACATCGATACCACATACCTGGTCCAGCTTATCTAG 909  
QY 363 ACATTCGATGCTTTTGAAGTCTCTATCAATTCGGAACAGATAT 408  
DB 908 CCTGATGCTTTTGAAGTCTCTATCAATTCGGAACAGATAT 863

## RESULT 9

US-08-633-770A-11/c

Sequence 11; Application US/08633770A  
Patent No. 5908760

## GENERAL INFORMATION:

APPLICANT: Bojsen, Kirsten  
APPLICANT: Yu, Shukun  
APPLICANT: Kragh, Karsten  
APPLICANT: Christensen, Torve  
APPLICANT: Marcussen, Jan  
TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,770A  
FILING DATE: July 8, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP94/03398  
FILING DATE: OCT-15-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: DY006.001APC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
TELEX:

## INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA



APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 520 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/04/056,220  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507...1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645...2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512...8070  
US-04-056-200-93

Query Match 2.2% Score 32.4; DB 1; Length 9551;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 81; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 3 CGGAACTCCATGTCGCCGTGATCGCTGGTCGAGCAGCATATTCGATGATCCAGCGGCC 62  
DB 3064 CGAGAGCCCAAGACAGAGTGTTCCAGGAGGAGACAGAGCTGAGGAGAGCCGAG 3123  
QY 63 GCACAGCGCTGTCGGGTATGATTAAGGCTGGTTGTAGAACCGCTGACCCCAACAG 122  
DB 3124 ACAGTCTCCGGAAGGAAGAGAGTTTCAGGAGAGAGAGAGCGGCGGCAAGAGAG 3183  
QY 123 CTCCTCATGATCTTTTAAATGCGCTGCTCTGCTCTGTAA 164  
DB 3184 CTCGAGGAGGAAGAGAGAGCTACGAGAGCTGGAGCGGCAA 3225

RESULT 13  
US-04-800-644-93  
Sequence 93, Application US/04/0644  
Patent No. 5958752  
GENERAL INFORMATION:  
APPLICANT: Steiner, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu

APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: F-dpsy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,644  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/056,220  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507...1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645...2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512...6070  
US-08-800-644-93

Query Match 2.2% Score 32.4; DB 2; Length 9551;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 3 CGGAACTCCATGTCGCCGTGATCGCTGGTCGAGCAGCATATTCGATGATCCAGCGGCC 62  
DB 3064 CGAGAGCCCAAGACAGAGTGTTCCAGGAGGAGACAGAGCTGAGGAGAGCCGAG 3123  
QY 63 GCACAGCGCTGTCGGGTATGATTAAGGCTGGTTGTAGAACCGCTGACCCCAACAG 122  
DB 3124 ACAGTCTCCGGAAGGAAGAGAGTTTCAGGAGAGAGAGAGCGGCGGCAAGAGAG 3183  
QY 123 CTCCTCATGATCTTTTAAATGCGCTGCTCTGCTCTGTAA 164  
DB 3184 CTCGAGGAGGAAGAGAGAGCTACGAGAGCTGGAGCGGCAA 3225

RESULT 14  
US-08-374-645-4/C  
Sequence 4, Application US/08374686  
Patent No. 5616474  
GENERAL INFORMATION:





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: July 6, 2003, 02:01:46 ; Search time 3753 Seconds  
(without alignments)  
11182.053 Million cell updates/sec

Title: US-10-086-082-1  
Perfect score: 1442  
Sequence: 1 cccgggaactccatgtggcc.....aatgcaatcatctggatcc 1442

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455:402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb.ba.\*  
2: gb.hlg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pi.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1442	100.0	1442	6	A69475 Sequence 1
2	1442	100.0	1442	6	A72152 Sequence 1
3	74	5.1	339650	1	AP003583 Nostoc sp
4	61.4	4.3	13698	1	AE005929 Caulobact
5	60.4	4.2	360150	1	AP001507 Bacillus
6	59.6	4.1	10485	1	AE006875 Sulfolobu
7	58.8	4.1	7983	1	AE013142 Thermoana
8	56	3.9	110000	2	Continuation (2 of
9	55.8	3.9	263050	1	AP000981 Sulfolobu
10	46.4	3.2	294800	1	AE000059 Aeropyrum
11	45.6	3.2	167300	1	AP000059 Sinorhizo
12	42.8	3.0	35346	8	SPAC869
13	41.2	2.9	307120	1	CNSPAX03
14	41.2	2.9	349061	1	NMA222491
15	41.2	2.9	349980	6	AX041919 Neisseria
16	40.8	2.8	716	11	PM11H4G
17	40.8	2.8	793	11	PM9D9G
18	39.6	2.7	92918	2	AC020050 Drosophill
19	39.6	2.7	130583	2	AC007420 Drosophill
20	39.6	2.7	155566	3	AC092231 Drosophill
21	39.6	2.7	170064	3	AC008323 Drosophill
22	39.6	2.7	304383	3	AE003579 Drosophill
23	39.4	2.7	118593	9	HS28811
24	39.2	2.7	1893	3	AY075188
25	39.2	2.7	10619	2	AF002546 Neisseria
26	39.2	2.7	133144	2	AC015424 Drosophill
27	39.2	2.7	150399	2	AC009369 Drosophill
28	39.2	2.7	183291	3	AC010003 Drosophill
29	39.2	2.7	183911	3	AC104507 Drosophill
30	39.2	2.7	271237	3	AE003521 Drosophill
31	39.2	2.7	345980	6	AX044034 Sequence
32	39.2	2.7	125620	9	AF429315 Homo sapi
33	38.4	2.7	10029	1	AE011694 Xanthomon
34	38.4	2.7	151085	8	AP003239 Oryza sat
35	37.6	2.6	503	6	I33795 Sequence 15
36	37.6	2.6	1233	6	AR153490 Sequence
37	37.6	2.6	1233	6	AR153491 Sequence
38	37.6	2.6	1233	6	AR153492 Sequence
39	37.6	2.6	3900	6	I33788 Sequence 6
40	37.6	2.6	3902	3	DMC31226 Drosophila
41	37.4	2.6	10029	1	AE009093 Agrobacte
42	37.4	2.6	14140	1	AE008058 Agrobacte
43	37.4	2.6	176688	2	AC084696 Mus muscu
44	37.4	2.6	162017	2	AL824710 Mus muscu
45	37.4	2.6	231853	2	AC093352 Mus muscu

ALIGNMENTS

RESULT 1  
A69475  
LOCUS  
DEFINITION Sequence 1 from Patent WO9802551.  
ACCESSION A69475  
VERSION A69475.1 GI:4760231  
KEYWORDS  
SOURCE  
ORGANISM Klebsiella oxytoca.  
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Klebsiella.

REFERENCE  
AUTHORS (bases 1 to 1442)  
Hovenberg, R.A., Van, D.L., Kerkman, R. and Nieboer, M.  
TITLE IMPROVED PROCESS FOR THE PRODUCTION OF ADIPOYL CEPHALOSPORINS  
JOURNAL Patent: WO 9802551-A 1 22-JAN-1998;

linear PAT 06-MAY-1999

Post dated

> post dated.

```
FEATURES
Source
G:ST BROCADES BV (NL)
Location/Qualifiers
1..1442
/organism="Klebsiella oxytoca"
/strain="PRS1"
/isolate="PRS1"
/db_xref="taxon:571"
197..1183
/codon_start=1
/transl_table=1
/product="AMIDASE"
/protein_id="CAB42489.1"
/db_xref="GI:4760232"
/translation="MKWLESIMAKRGVGPVTHLTEDMKFFHYTIGPSTPV
LIIEPDGRIIVDRDAPEAINSEODIPGOLKPELNPONGPIVYKAGKGDVAVY
IESMLRGVDPYICIGAMIPHEGLGTGDTILAMLDNPEPKVRMIKLSKVIWNRHT
LPKPHITGLSVFEDSINSINLPDRHGNMVPDIPGUSITYVPRAPGGRLFGGA
HACGGDEIGSTAVEFASITIKVDLIKWLQSKPRMENAEINIMTSGSRPIFDA*RI
AYRDLIYLVWVED*GFQWDAYMILISQKVRIGNMYDPKTYTGAME.NKNELV"
BASE COUNT 385 a 350 c 370 g 337 t
ORIGIN
Query Match 100.0%; Score 1442; DB 6; Length 1442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGGGAACTCCATGTGCGCGTGAICTGTGTCGAGCAGGATATTGGGATGATCCAGCGG 60
DB 1 CCGGGAACTCCATGTGCGCGTGAICTGTGTCGAGCAGGATATTGGGATGATCCAGCGG 60
QY 61 CCGCACAGCGCTGCGGTGAATGATGAATAAGCGCTGGTTHAGAAACCGCGACCCACAC 120
DB 61 CCGCACAGCGCTGCGGTGAATGATGAATAAGCGCTGGTTHAGAAACCGCGACCCACAC 120
QY 121 AGCTCTGTGATGATCTTTAAAGCGCTGATCTGCTGCTGTAACTAAAGCGCTAAAT 180
DB 121 AGCTCTGTGATGATCTTTAAAGCGCTGATCTGCTGCTGTAACTAAAGCGCTAAAT 180
QY 181 ACGTGGAGATAACATATGAATAGTATGTTGAGAGATTCATATGGCGAAAGCGGTG 240
DB 181 ACGTGGAGATAACATATGAATAGTATGTTGAGAGATTCATATGGCGAAAGCGGTG 240
QY 241 TGGCGGCGTAAACCGGTACCGATACCTGACGCGGAGAAATGCAAAAGAGTTCATTA 300
DB 241 TGGCGGCGTAAACCGGTACCGATACCTGACGCGGAGAAATGCAAAAGAGTTCATTA 300
QY 301 CACCATTTGGCCCTTATTCACACCGCTGCTGACCATCGACCGGCTGACCGGATATGT 360
DB 301 CACCATTTGGCCCTTATTCACACCGCTGCTGACCATCGACCGGCTGACCGGATATGT 360
QY 361 CGACACTCGAGATGCTTTTGAAGTGTCTATCAATTCGGAACAGGATATTCGAGCGAGT 420
DB 361 CGACACTCGAGATGCTTTTGAAGTGTCTATCAATTCGGAACAGGATATTCGAGCGAGT 420
QY 421 GCTAAAAATGCCCTTTCTCAACCCACAAACGACCGATCATGCTCAATGGCGCGAGAA 480
DB 421 GCTAAAAATGCCCTTTCTCAACCCACAAACGACCGATCATGCTCAATGGCGCGAGAA 480
QY 481 AGGTGATGTCGCTGCTATATCGAATCCATGTTGCGCGCGGCTGTCATCCCTACGG 540
DB 481 AGGTGATGTCGCTGCTATATCGAATCCATGTTGCGCGCGGCTGTCATCCCTACGG 540
QY 541 CATCTGCGCATGATCCGATTTTGGCGGACTGACGCGGACCGGCTGACGCGCATGCT 600
DB 541 CATCTGCGCATGATCCGATTTTGGCGGACTGACGCGGACCGGCTGACGCGCATGCT 600
QY 601 CAATGATCCGCTGCCAGAAAGGTGGCGCATGATTAACCTCGACAGTGAAGAAGTCTACTG 660
DB 601 CAATGATCCGCTGCCAGAAAGGTGGCGCATGATTAACCTCGACAGTGAAGAAGTCTACTG 660
QY 661 GAGCAACGCGATACGCTTCCCTATTAACCCCGCATATGCGACCGCTGAGCGCATGCGAGA 720
DB 661 GAGCAACGCGATACGCTTCCCTATTAACCCCGCATATGCGACCGCTGAGCGCATGCGAGA 720
```

## RESULT: 2

A72152

LOCUS

DEFINITION

A72152

ACCESSION

A72152.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

A72152 Sequence 1 from Patent WO9801568. 1442 bp DNA linear PAT 11-MAY-1999

A72152.1 GI:4808107

Klebsiella oxytoca.

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Klebsiella.

1 (bases 1 to 1442)

Brieden, W., Naughton, A., Robins, K., Shaw, N., Tinschert, A. and

Zimmermann, T.

METHOD OF PREPARING (S) - OR (R) - 3,3,3-TRIFLUORO-2-HYDROXY-2-

METHYLPROPIONIC ACID

Patent: WO 9801566-A 15-JAN-1998;

LOCUS AG (CH)

Location/Qualifiers

applicant's own seq. 100% identical

```
source 1. .1442
/organism="Xiebiella oxyloca"
/strain="PRSI"
/isolate="PRSI"
/db_xref="taxon:57"
197..1183
CDS
/codon_start=1
/transl_table=11
/product="AMIDASE"
/protein_id="CAB42722.1"
/db_xref="GI:4808108"
/translation="MKWLEESIMAKRGVCAKPKVTHHITTEMKHFHVTIGVSTPV
IEMPRGVDPYGCAMIPKPKKIDSEKLVYWSKRV
LPHKPHGTVSPIDNSLTPDHGNSXDPDIFGSIYLPVAPGKELFIDGA
HACQDGEICGTAFVAFSTITIKVDLKNWLSWPMENAEINMISGARPLEDATE
AYRDIYMLVDFGFEOMDAYMLISOGCKVRGNKVTPKYTVGAMLNKLLV"
BASE COUNT 385 a 350 c 370 g 337 t
ORIGIN
Query Match: 100.0% Score 1442; DB 6; Length 1442;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGGGAAGTCATGTGGCGGTGATCCTGGTGGACACAGATATTCGGATGATCCAGCGG 60
DB 1 CCGGGAAGTCATGTGGCGGTGATCCTGGTGGACACAGATATTCGGATGATCCAGCGG 60
QY 61 CCGCACAGCGGTGCGGTAAATGGATAAAGGCGTGTGTAGAAAGCGTGACCGCAACAC 120
DB 61 CCGCACAGCGGTGCGGTAAATGGATAAAGGCGTGTGTAGAAAGCGTGACCGCAACAC 120
QY 121 AGCTCTCTGATGATCTTTAAATGGCTGTCATCTGGCTGTGAACCTAAACGCTATAAAT 180
DB 121 AGCTCTCTGATGATCTTTAAATGGCTGTCATCTGGCTGTGAACCTAAACGCTATAAAT 180
QY 181 AGCTGGAGAAATACATATCAATGTTTGAAGAATCCATTATGGCCAAACGCGGTGG 240
DB 181 AGCTGGAGAAATACATATCAATGTTTGAAGAATCCATTATGGCCAAACGCGGTGG 240
QY 241 TCGCGGCGGTAAACCGGTAAACGCATCACCCTGACGGAAGAAATGCCAAAGAGTTTCAAT 300
DB 241 TCGCGGCGGTAAACCGGTAAACGCATCACCCTGACGGAAGAAATGCCAAAGAGTTTCAAT 300
QY 301 CACCATGGCCCTTATTCACACCGCTGCTGACATCGAACCGGTGACCAATTAATGT 360
DB 301 CACCATGGCCCTTATTCACACCGCTGCTGACATCGAACCGGTGACCAATTAATGT 360
QY 361 CGACACTCGAGATGCTTTTGAAGGTGCTATCAATTCGGAACAGGATATTCGACGCGATT 420
DB 361 CGACACTCGAGATGCTTTTGAAGGTGCTATCAATTCGGAACAGGATATTCGACGCGATT 420
QY 421 GCTAAATATGCGCTTTCTCAACCCACAAACGACCGATGATGTCATGTCGCGAGAA 480
DB 421 GCTAAATATGCGCTTTCTCAACCCACAAACGACCGATGATGTCATGTCGCGAGAA 480
QY 481 AGGIGATGCTGGTGTCTATATCAATCCATGTTGCGCGCGGTGATGTCGCTACG 540
DB 481 AGGIGATGCTGGTGTCTATATCAATCCATGTTGCGCGCGGTGATGTCGCTACG 540
QY 541 CATCTCGGCATGATTCGCGATTTTGGCGAGTACCGGAGGAGACCTGACGCGCATGTT 600
DB 541 CATCTCGGCATGATTCGCGATTTTGGCGAGTACCGGAGGAGACCTGACGCGCATGTT 600
QY 601 CAATGATCGCTGCCAGAAAGGTGGCGATGATTAACCTCGACAGTGAAGAAGTCTAC 660
DB 601 CAATGATCGCTGCCAGAAAGGTGGCGATGATTAACCTCGACAGTGAAGAAGTCTAC 660
QY 661 GAGCAACGCCATACGCTTCCCTATTAACCCCATATTCGCTGAGCGTATCCGCGA 720
DB 661 GAGCAACGCCATACGCTTCCCTATTAACCCCATATTCGCTGAGCGTATCCGCGA 720
QY 721 AATTGACTCAATCAATTCATCGACGCCAGACAATACCGCGGGAATATGATGTGGCGA 780
```

```
DB 721 AATTGACTCAATCAATTCATCGACGCCAGACAATACCGCGGGAATATGGAATGCCGGA 780
QY 781 TATAGGACACGGAGATTAACCTATCTCCCGGTAGTGGCGCTGGAGCGCGCTGTTAT 840
DB 781 TATAGGACACGGAGATTAACCTATCTCCCGGTAGTGGCGCTGGAGCGCGCTGTTAT 840
QY 841 TSGTGATSCCATGCTTGTACAGGGTGTATGAGATTTGCGGGACCGCAGTAGAGTTGC 900
DB 841 TSGTGATSCCATGCTTGTACAGGGTGTATGAGATTTGCGGGACCGCAGTAGAGTTGC 900
QY 901 CTCGAATCACCACCAACAAAGTCGATTTGATCAAACTGSCAGCTTCCCTGGCCACGAAT 960
DB 901 CTCGAATCACCACCAACAAAGTCGATTTGATCAAACTGSCAGCTTCCCTGGCCACGAAT 960
QY 961 GGAGAAATGCCGAAATATATGAGTATTTGCACTGTCAGCTCCGCTGGAGGATCGAGCG 1020
DB 961 GGAGAAATGCCGAAATATATGAGTATTTGCACTGTCAGCTCCGCTGGAGGATCGAGCG 1020
QY 1021 AATTGCAATATCGGACITTAATTTACTTGGTGGTGAAGACTTTTGGCTTGAACAAATGGA 1080
DB 1021 AATTGCAATATCGGACITTAATTTACTTGGTGGTGAAGACTTTTGGCTTGAACAAATGGA 1080
QY 1081 TGCCTACATGCTTCTGAGTCAATCGGCAAGTGGGCAACATGGTGGACCCCAA 1140
DB 1081 TGCCTACATGCTTCTGAGTCAATCGGCAAGTGGGCAACATGGTGGACCCCAA 1140
QY 1141 ATACACCGTTTGGCGGATGCTGTAACAAACAACTTTAGTATAGGAATAAATTAAC 1200
DB 1141 ATACACCGTTTGGCGGATGCTGTAACAAACAACTTTAGTATAGGAATAAATTAAC 1200
QY 1201 TGAACATATACCGGATGCTGTAACAAACAACTTTAGTATAGGAATAAATTAAC 1260
DB 1201 TGAACATATACCGGATGCTGTAACAAACAACTTTAGTATAGGAATAAATTAAC 1260
QY 1261 TGCCTACATGCTTCTGAGTCAATCGGCAAGTGGGCAACATGGTGGACCCCAA 1320
DB 1261 TGCCTACATGCTTCTGAGTCAATCGGCAAGTGGGCAACATGGTGGACCCCAA 1320
QY 1321 TTCAGCGAGTGGCGGCGACACAGAGTGTCAACATGGCGGAGCAACCCAGGCTATT 1380
DB 1321 TTCAGCGAGTGGCGGCGACACAGAGTGTCAACATGGCGGAGCAACCCAGGCTATT 1380
QY 1381 GCGCAATTAATCAAAATGCGGCGATCAACCGGAGACCACTCAATGCAATTCATTTGAT 1440
DB 1381 GCGCAATTAATCAAAATGCGGCGATCAACCGGAGACCACTCAATGCAATTCATTTGAT 1440
QY 1441 CC 1442
DB 1441 CC 1442
RESIDUE 3
AP003583/3 339650 bp DNA linear BCT 28-NOV-2001
DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section 3/19.
ACCESSION AP003583.1
VERSION AP003583.1 G:17129939
KEYWORDS Nostoc sp. PCC 7120 DNA.
SOURCE Nostoc sp. PCC 7120
ORGANISM Nostoc sp. PCC 7120
Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE 1
AUTHORS Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S.,
Watanabe, A., Iriuchimi, M., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kofara, M., Matsumoto, M., Matsuno, A., Muraki, A.,
Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yamada, M.,
Yasuda, M. and Tabata, S.
Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120
JOURNAL DNA Res. 8 (5), 205-213 (2001)
MEDLINE 21595285
REFERENCE 2 (bases 1 to 339650)
```

# AUTHORS

kaneko, T.  
 Direct Submission  
 Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research, 1-ana  
 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: kaneko@kazusa.or.jp,  
 URL: http://www.kazusa.or.jp/cyanobase/  
 Tel: 81-438-52-3935(ex. 2336), Fax: 81-438-52-3934)

## FEATURES

```

source
1..339650
  /organism="Nostoc sp. PCC 7120"
  /db_xref="taxon:103690"
  /note="synonym: Anabaena sp. PCC7120"
  /complement(98..2698)
  /gene="all0594"
  /complement(98..2698)
  /gene="all0594"
  /note="ORF_ID:all0594"
  /codon_start=1
  /transl_table=11
  /product="aminopeptidase"
  /protein_id="BAB72552.1"
  /db_xref="GI:17129940"
  /translation="MPHSYFDDQSNNAHKSFEFGARPHYNPDPQOVHEHIFLDLND
  IPOQYEGSGSIRLLPIRNGIDRLTLDVNLQJRSVQVDEVGQDFDYDGLVLIQLSQ
  PTEIGKRILIAIYAVEKPGQGIYFIQPDSHYPNKTQVMTQGEDESRWFPCFQVY
  GQLSTSEIRVRPKSLMAISNGELIDIOEDGKHKIYHWSQOQIHPTVYLM-LAVADYAF
  IRDWRGKPTVYVYVEKGRADAKRSMGKTPIRIFFELSEKYGVAAPKPKACQVDSFI
  FGGHNTSTLLIDRCLLDERAALDNKTESLVVHELHQWFGDLIVIKWASHAKIKE
  QMASYSEVMTEQDISAEAAAYRIILSARSTLSDSGKRRPVMVHYIREAFLYDHH
  CYEYRGYHMI:RAELGEDLPQWA-QTFVRDNAHKTVEI:DLLEAIEKATGRNLAF:LF
  DQYVRSCHPDFKAYVSDWGHAKLAKVTVTQVPAKDQKDFDLRIPIFGVGNQ:EA
  AQSITQTVRVHEQSEFYPLETEKPDF-SFDVGNHMLKTVSLSEYPIPELKAQIASGP
  DPISIIYAAALAKKGLGKATKA:SAALKNEPWPVGRVEKALAEIKILQAGLQVE
  GLADKNAYVRRAVQAALAKTKTLASRYAKGVYKDGDSYVESAAALGCA:NAIWF
  DLRKAKAKLILOSLEERAGWNEVRSAGIGLAELK:SEAAALDL:IHYIKIGVYQ
  PDKAAIRAGKATSPGSPGNLERIIARJAELESKENFFLTIMAYVAALGMQETPKAG
  VLRLAQDTAGRVYRFAEELANVQXNIGTGNALRLKLRFFCDLKQCNQELKSRLEN
  LEAKSQTVKQ"
  complement(2862..3059)
  /gene="as10595"
  /complement(2862..3059)
  /note="ORF_ID:as10595"
  similar to cytochrome c oxidase chain I"
  /codon_start=1
  /transl_table=11
  /protein_id="BAB72553.1"
  /db_xref="GI:17129942"
  /translation="MRHCVHP:IIHLS:FLYKIIILN:QNKPSDRPHKVGILYYAFRYK
  IAIQFF:AGLMVLSTLFFEVF"
  complement(3151..3810)
  /gene="all0596"
  /complement(3151..3810)
  /note="all0596"
  /note="ORF_ID:all0596"
  /codon_start=1
  /transl_table=11
  /protein_id="BAB72554.1"
  /db_xref="GI:17129942"
  /translation="MGVCHSKDVSEFC:JGRFMFEVFKDGYKLGKLLLTADGECYVK
  LSKQRFDFWRLPTGLLQVYGEKKYTKLAEAVMAAATETSSAVKQVPM
  ENTPKPKAKATILVQKDCMRGKALKQALEATLSDRGLQDVYIKGTGCMKNC
  KAGPNLVMPDKTRYTRICADQVPKLMNKHFDASQSDSHSQAVISINTKPEPFEVSA
  F"
  complement(3877..4134)
  /gene="as10597"
  /complement(3877..4134)
  /gene="as10597"
  /note="ORF_ID:as10597"
  /codon_start=1
  /hypothetical protein"

```

## gene

```

CDS
1..5146
  /transl_table=11
  /protein_id="BAB72555.1"
  /db_xref="GI:17129943"
  /translation="MKQFSSEVEEKSVVEVDADRQWYVQRIQLELPCWCEANQPLK
  VELVTPMTVVQLWSVMRQFTSSRQDLIYSLINMQSRYPYS"
  4451..5146
  /gene="alr0598"
  4451..5146
  /gene="alr0598"
  /note="ORF_ID:alr0598"
  /codon_start=1
  /transl_table=11
  /protein_id="BAB72556.1"
  /db_xref="GI:17129944"
  /translation="MYRTWRPLLSTLSVLVLQGGFPVIVAOGSENTGDVASTAIQOV
  IAAIIMTNSADGNFYPERLMNRAELASILVKAFRLNKOBAANOPKSIITVADYPRSHWA
  YODIOTVLTVDNMKGYNLFFPNORVTRAEALIPAFQAYGVFOFSDYVNEILAPYP
  DANSIPNARRAIATVITEGFINTDAQNNITPLRPMTRGDMAYLILSKYLQROOKOPDT
  PVVPTTIDSPQLP"
  5596..7503
  /gene="alr0599"
  5596..7503
  /note="ORF_ID:alr0599"
  /codon_start=1
  /transl_table=11
  /product="1-deoxy-xylulose 5-phosphate synthase"
  /protein_id="BAB72557.1"
  /db_xref="GI:17129945"
  /translation="MHLSEITHPNQLHGLSIRLOQIAKQIRDKHLQTVAEFGHGLGP
  GLGVELTGLYQTDLDLRQKVTVDVGHQAYPHKLITGRYSDFHTLRQKDGIAGYLKR
  GENKDFHGCAGHASTSI SAALGNALARDMNGEKFAVAVIGDCAITGGMALPAINHAG
  HLZKINLLVNLNDNEMSI PNVCAGI PRYLNNKMLSPVOFIDKNFEQFKHIPPVGES
  LSPLEIRKEMGRALAVPVGVGVEELGTITMGVDPGHNELEIATFQAHQHPAGPVL
  VHVATIKGGLYELAEQDKQVYHAQTFNLTITGAIPNKPYPAYAKVFSHTLVKLAIE
  QNPK1IGITAAATAGTGLDKLQALPNQYIDVGAIEQHAVTLAAGLATETGMRVAAIY
  STFLQRAYDQI:HDVCIONLPVFCLDRAGIVGSDGPTHQGYDIDYALICPINVMA
  FKPSAEQRMVVTGIEHTTGPIDAMPFRPGNGYGVPLMEEGWPLEIGKGEILRNGDV
  LIIGYGMYPSPNOAREILSEHGLEATVINARFKPLDTLIVPLAQKIGRVVTLLEG
  CYMKGGSVAVEALLDADVVPVVRIGIPDVLVEHAIPDESKAELGLTSRQIAERVIO
  AYFQKQSPVS"
  7793..8281
  /gene="alr0600"
  7793..8281
  /gene="alr0600"
  /note="ORF_ID:alr0600"
  /codon_start=1
  /transl_table=11
  /protein_id="BAB72558.1"
  /db_xref="GI:17129946"
  /translation="MADELKVKIECFILSKLSMKLNFSMKKLSNLGRWIAITTVFCL
  SAIAFYWGGAFFADTSAMADNSTYLIATSDATSLRNKSKEFIRDTQKNVKAARTNA
  SRVYDARDNGSTAEKANKASRIQRAEEDARTETAYVDNNLNAVERTIDNIKGAFT
  NN"
  complement(8399..9469)
  /gene="all0601"
  /complement(8399..9469)
  /gene="all0601"
  /note="ORF_ID:all0601"
  /hypothetical protein"
  /codon_start=1
  /transl_table=11
  /protein_id="BAB72559.1"
  /db_xref="GI:17129947"
  /translation="NSIVFRDLLKVGSGNHTGESLNRYEAATATKMMFLGEATPAQI
  GAFLIAHRKRPTEGEAGMLDAFDELGPKLQPISSKRPIVIFGIPYDGRTRTATLSP
  VTALLMATVEQPVYIHGGDRLPTKYGLPLTEIMQGLGVNMTLSLEKTEQVEAQTGIG
  FVYTPKHFTLINSIWEYRQDLGKRPPLATWELIWCPCYAGDAHVHYHPPTGEMFQV
  ALERGVSRFTLVKLEGSCDLPRPRTATISLSSAPELERVFLSTHYTGFTIKNVP
  GSTEELLTQIDVLAGKLGLOQTALWNGFGYLRWSGICADMRSGIAKAEELNSGLV
  AAKLQELIQVNSVTENLLSIT"

```

## gene

```

CDS
1..5146
  /transl_table=11
  /protein_id="BAB72555.1"
  /db_xref="GI:17129943"
  /translation="MKQFSSEVEEKSVVEVDADRQWYVQRIQLELPCWCEANQPLK
  VELVTPMTVVQLWSVMRQFTSSRQDLIYSLINMQSRYPYS"
  4451..5146
  /gene="alr0598"
  4451..5146
  /gene="alr0598"
  /note="ORF_ID:alr0598"
  /codon_start=1
  /transl_table=11
  /protein_id="BAB72556.1"
  /db_xref="GI:17129944"
  /translation="MYRTWRPLLSTLSVLVLQGGFPVIVAOGSENTGDVASTAIQOV
  IAAIIMTNSADGNFYPERLMNRAELASILVKAFRLNKOBAANOPKSIITVADYPRSHWA
  YODIOTVLTVDNMKGYNLFFPNORVTRAEALIPAFQAYGVFOFSDYVNEILAPYP
  DANSIPNARRAIATVITEGFINTDAQNNITPLRPMTRGDMAYLILSKYLQROOKOPDT
  PVVPTTIDSPQLP"
  5596..7503
  /gene="alr0599"
  5596..7503
  /note="ORF_ID:alr0599"
  /codon_start=1
  /transl_table=11
  /product="1-deoxy-xylulose 5-phosphate synthase"
  /protein_id="BAB72557.1"
  /db_xref="GI:17129945"
  /translation="MHLSEITHPNQLHGLSIRLOQIAKQIRDKHLQTVAEFGHGLGP
  GLGVELTGLYQTDLDLRQKVTVDVGHQAYPHKLITGRYSDFHTLRQKDGIAGYLKR
  GENKDFHGCAGHASTSI SAALGNALARDMNGEKFAVAVIGDCAITGGMALPAINHAG
  HLZKINLLVNLNDNEMSI PNVCAGI PRYLNNKMLSPVOFIDKNFEQFKHIPPVGES
  LSPLEIRKEMGRALAVPVGVGVEELGTITMGVDPGHNELEIATFQAHQHPAGPVL
  VHVATIKGGLYELAEQDKQVYHAQTFNLTITGAIPNKPYPAYAKVFSHTLVKLAIE
  QNPK1IGITAAATAGTGLDKLQALPNQYIDVGAIEQHAVTLAAGLATETGMRVAAIY
  STFLQRAYDQI:HDVCIONLPVFCLDRAGIVGSDGPTHQGYDIDYALICPINVMA
  FKPSAEQRMVVTGIEHTTGPIDAMPFRPGNGYGVPLMEEGWPLEIGKGEILRNGDV
  LIIGYGMYPSPNOAREILSEHGLEATVINARFKPLDTLIVPLAQKIGRVVTLLEG
  CYMKGGSVAVEALLDADVVPVVRIGIPDVLVEHAIPDESKAELGLTSRQIAERVIO
  AYFQKQSPVS"
  7793..8281
  /gene="alr0600"
  7793..8281
  /gene="alr0600"
  /note="ORF_ID:alr0600"
  /codon_start=1
  /transl_table=11
  /protein_id="BAB72558.1"
  /db_xref="GI:17129946"
  /translation="MADELKVKIECFILSKLSMKLNFSMKKLSNLGRWIAITTVFCL
  SAIAFYWGGAFFADTSAMADNSTYLIATSDATSLRNKSKEFIRDTQKNVKAARTNA
  SRVYDARDNGSTAEKANKASRIQRAEEDARTETAYVDNNLNAVERTIDNIKGAFT
  NN"
  complement(8399..9469)
  /gene="all0601"
  /complement(8399..9469)
  /gene="all0601"
  /note="ORF_ID:all0601"
  /hypothetical protein"
  /codon_start=1
  /transl_table=11
  /protein_id="BAB72559.1"
  /db_xref="GI:17129947"
  /translation="NSIVFRDLLKVGSGNHTGESLNRYEAATATKMMFLGEATPAQI
  GAFLIAHRKRPTEGEAGMLDAFDELGPKLQPISSKRPIVIFGIPYDGRTRTATLSP
  VTALLMATVEQPVYIHGGDRLPTKYGLPLTEIMQGLGVNMTLSLEKTEQVEAQTGIG
  FVYTPKHFTLINSIWEYRQDLGKRPPLATWELIWCPCYAGDAHVHYHPPTGEMFQV
  ALERGVSRFTLVKLEGSCDLPRPRTATISLSSAPELERVFLSTHYTGFTIKNVP
  GSTEELLTQIDVLAGKLGLOQTALWNGFGYLRWSGICADMRSGIAKAEELNSGLV
  AAKLQELIQVNSVTENLLSIT"

```





CY	326 TTGATCAGAA 936	REFERENCE	9 (sites)
Db	11 11111	AUTHORS	Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T., Hirama,C., Fujii,F. and Takami,H.
	438 CTGCACAAGAA 428	TITLE	Characterization and comparative study of the rrr operons of alkaliphilic <i>Bacillus halodurans</i> C-125
RESULT 5		JOURNAL	<i>Extremophiles</i> 4 (4), 209-214 (2000)
AP001507/c		MEDLINE	20426005
LOCUS	AP001507 300150 bp DNA linear BCT 10-JAN-2001	PUBMED	10972189
DEFINITION	<i>Bacillus halodurans</i> genomic DNA, section 1/14.	REFERENCE	10 (sites)
ACCESSION	AP001507 BA000304	AUTHORS	Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N., Fujii,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and Horikoshi,K.
VERSION	AP001507.1 GI:10172612	TITLE	Complete genome sequence of the alkaliphilic bacterium <i>Bacillus halodurans</i> and genomic sequence comparison with <i>Bacillus subtilis</i>
KEYWORDS	<i>Bacillus halodurans</i> DNA.	JOURNAL	<i>Nucleic Acids Res.</i> 28 (21), 4317-4331 (2000)
ORGANISM	<i>Bacillus halodurans</i>	MEDLINE	20512582
SOURCE	Bacteria: Firmicutes: Bacillales: Bacillaceae: <i>Bacillus</i> .	PUBMED	11058132
REFERENCE	1 (sites)	REFERENCE	11 (bases 1 to 300150)
AUTHORS	Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fujii,F., Nakamura,Y. and Inoue,A.	AUTHORS	Takami,H. and Takaki,Y.
TITLE	An improved physical and genetic map of the genome of alkaliphilic <i>Bacillus</i> sp. C-125	TITLE	Direct Submission
JOURNAL	<i>Extremophiles</i> 3 (1), 21-28 (1999)	JOURNAL	Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group; 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan
MEDLINE	99184645	REFERENCE	(E-mail:takami@jamstec.go.jp, URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html, Tel:81-468-67-3895, Fax:81-468-66-6364)
PUBMED	10086841	FEATURES	Location/Qualifiers
REFERENCE	2 (sites)	Source	1. 300150
AUTHORS	Takami,H. and Horikoshi,K.		/organism="Bacillus halodurans"
TITLE	Reidentification of facultatively alkaliphilic <i>Bacillus</i> sp. C-125 to <i>Bacillus halodurans</i>	gene	/db_xref="taxon:86665"
JOURNAL	<i>Biosci. Biotechnol. Biochem.</i> 63, 943-945 (1999)	CDS	584..1933
REFERENCE	3 (sites)		/note="alkaliphilie"
AUTHORS	Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y., Masui,N., Fujii,F., Inoue,A. and Horikoshi,K.		/gene="dnaA"
TITLE	Sequencing of three lambda clones from the genome of alkaliphilic <i>Bacillus</i> sp. strain C-125		584..1933
JOURNAL	<i>Extremophiles</i> 3 (1), 29-34 (1999)		/note="BH0001"
MEDLINE	99184646		/codon_start=1
PUBMED	10086842		/trans_table=1
REFERENCE	4 (sites)		/product="initiation of chromosome replication"
AUTHORS	Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and Horikoshi,K.		/protein_id="BAB03720.1"
TITLE	Sequence analysis of a 32-Kb region including the major ribosomal protein gene clusters from alkaliphilic <i>Bacillus</i> sp. strain C-125		/db_xref="GI:10172613"
JOURNAL	<i>Biosci. Biotechnol. Biochem.</i> 63 (2), 452-455 (1999)		/translations="MENIDHLEALKSMEKVKSPFETWLKUTKANSIEDSTLIIT APNEFFARDLEKHYDELISETIDLTGVRLYPKFVPISTQDEFFVEQLKPKMKOPP AONGEMNNLNDKYDTFTVIGSGNRFHAASLAAVAEAPAKYNFLFYGGVGLGKT ALHMAIGHVYMDHNPAKVYLSSEKFTNEFINAIRDNKAVNFRNRYRNVDVLLIDD1 QFLAGKQTOEEFFHTFNALNDKNQIVISSDRPPKEIPTLEDRLSRPFEGWGLTDT PPDETFTAILRLKKAENLDIPNEVMVLIANOIDNIRELGALIRVAVYSSLIIND MNAADLAFAUKDIIIPNAKPRVLTIDTOKTVGEYFHVKLEDFKAKRTKSVAFPROIA MYLSRELTASLPKIGSEFGGRDHTTVIHAHEKISKLLSTDQELQKIQDISKLSR"
MEDLINE	99209008	gene	2103..3245
PUBMED	10192928	CDS	2103..3245
REFERENCE	5 (sites)		/note="dnaA"
AUTHORS	Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G., Sasaki,R., Hirama,C., Fujii,F. and Masui,N.		/gene="dnaA"
TITLE	Genetic analysis of the chromosome of alkaliphilic <i>Bacillus halodurans</i> C-125		/note="BH0002"
JOURNAL	<i>Extremophiles</i> 3 (3), 227-233 (1999)		/codon_start=1
MEDLINE	99411980		/trans_table=1
PUBMED	10484179		/product="DNA polymerase III (beta subunit)"
REFERENCE	6 (sites)		/protein_id="BAB03721.1"
AUTHORS	Takami,H.		/db_xref="GI:10172614"
TITLE	Genome analysis of facultatively alkaliphilic <i>Bacillus halodurans</i> C-125		/translations="MHEVIDRDIFVQNVNHNVSXAVSRTTPIILTGIKIIVADHEGVTL TGSQDISIETFTPLLEEDGRQNVKQEGSIVLQAKVFAETVKKLPEQETEIHVQDSF VTIIRKSSVYENLGLDPPYPRLPVLEEDHVERLPQKTLKDLIRQTVFAVSTQETRP VLTGVNFEIDGILCTCTATDSHRLAMKRVKVRKNDELQFSNVV1PKSLSNELSKTID ENEELDIIYVTFNQTFLKLNMLFFSLRLEGKYPVTKNM1PKRAKTSFAVHTKAFIQ1 IFRALLSKREGKQVINLKTLDGVVEVTAITPEIGKVTENVAIQEGEELRSTNG KNVIDALKVYDSESIHIAFTGAMSPFVLSPTDHDQSLHLFSFVRVY"
JOURNAL	(in) <i>Extremophiles</i> in deep-sea environments (Ed.): HORIKOSHI, K. TSGJII: 249-284; Springer-Verlag (1999)	gene	3429..3650
MEDLINE	99356711	CDS	3429..3650
PUBMED	10427704		/note="BH0003"
REFERENCE	7 (sites)		/gene="BH0003"
AUTHORS	Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.		3429..3650
TITLE	Replication origin region of the chromosome of alkaliphilic <i>Bacillus halodurans</i> C-125		/note="BH0003"
JOURNAL	<i>Biosci. Biotechnol. Biochem.</i> 63 (5), 1134-1137 (1999)		unknown conserved protein"
MEDLINE	99356711		/codon_start=1
PUBMED	10427704		
REFERENCE	8 (sites)		
AUTHORS	Takami,H. and Horikoshi,K.		
TITLE	Analysis of the genome of an alkaliphilic <i>Bacillus</i> strain from an industrial point of view		
JOURNAL	<i>Extremophiles</i> 4 (2), 99-108 (2000)		
MEDLINE	20263314		
PUBMED	10805564		









```
/translation--"MKFFELTANIESIKER":SLGVISGVTTNPSLVAKEGRJFKEVIR
EIAEVDGPISAEVI:SDHECMYKEARELAKIHKNIVIRIPMTAFGRKAVNIIISKEGI
KTAVTITFSANKAZLAARAGATVYSPFGR:LDINDGMOITFCIVTIL:TNVDIOARI
ITASVRHPIHVI:FAAKLGAHIATVP:KVIHQMIK:HLTDIGISERKEHKKACLK:
complement(3193..3637)
/gene="MipB"
/note="Pfam match to entry Transaldolase, Transaldolase,
score 311.1, E-value 3.40E-92"
complement(3851..3855)
/gene="MipB"
3984..4896
/gene="TTE1919"
3984..3988
/gene="TTE1919"
3997..4896
/gene="TTE1919"
/note="Best Blastp hit - g1110172637 dbj|BAU03744.1|
(AF001507) acetamidase [Bacillus halodurans], score 202,
E-value 3.00E-51"
/codon_start=1;
/transl_table=1;
/product="predicted acetamidase/formamidase"
/protein_id="AAK25099.1"
/db_xref="GI:20516939"
/translation="MKYSIACDHIIFAFSKENKPAISVXSGDLEVEYTCGCSNLIQS
NEQKLDPMNHNVPALGP:FEVSAKEGVKIKK:IEVAEKGVLAIGKGLVNLN:
MEGLYSKVBIKGGVIFNEKALPKNPTGVIGVAPKEGSGNCGP:SGHGGKNDII:
IAEGAEYFVFVEGALLAIGDILHMGDGEVGVGVAGVKGVLLEVEVIRKGN:KRP
VKTAEYATIASAESLDKAVEIAVHDMAELFKRHTDLSGEGATLFTSGTGNAGISQV
VDELKTARFSLPNWIILESYGRF"
4896..5602
/gene="CcmA10"
/note="CcmA10"
4896..4900
/gene="CcmA10"
4910..5602
/gene="CcmA10"
/note="Best Blastp hit - g113814525|gb|AAK41557.1.
(AE006746) Antibiotic ABC transporter, ATP binding
protein [Salifolobus solifarius], score 102, E-value
6.00E-21"
/codon_start=1;
/transl_table=1;
/product="ABC-type multidrug transport system, ATPase
component"
/protein_id="AAK25100.1"
/db_xref="GI:20516930"
/translation="MIAEEIVKLENTVKLGRLTILEDINFEVKEGKVKILGPKKSG
KTALVEITLGI:KPDGVSVLGTDPTYNSSSL:GAASVYKNSFPLHWRVKNQVNF
KKOLWGENSOHLIFALDISNMKNTTVGKLSERAKLNLFFALTEKALITLGDPA
NSLNFDLENIKNPHTSFTIITSKTPPVYDIADRIYILHEGRIKKSIAKQDYSSLEK
LKS:LYOYAPED"
5003..5515
/gene="CcmA10"
/note="Pfam match to entry ABC_tran, ABC transporter,
score 23.5, E-value 3.53E-07"
5586..6287
/gene="TTE1921"
5586..6287
/gene="TTE1921"
/transl_table=1;
/product="hypothetical protein"
/protein_id="AAK25101.1"
/db_xref="GI:20516931"
/translation="MRUKTYGVYEVKIRADALKVFTTISSEIRFSSIPQVPL:
LLGYAGAGKGYIPDEF:LWKLMSGMALYIATIPWGAALRERQSGNIR:MLSG
VSPHEW:GKSFLLSGISYIALISPLGYKINPTWOLVPEWAP:FLGFFIIIS
LTLFLCAVGLSVAGTSMAFTISYTLFGLFGYISPKRSSEILILFLFLFLF:KSF
LVAKRPEKEMAYI"
5291..6316
/note="putative rho-independent transcription terminator"
complement(6324..7776)
```

```
/gene="PotB3"
/note="TTE1922"
complement(6324..7763)
/gene="PotB3"
/note="Best Blastp hit - g116759463|emb|CAB69791.1|
(AJ237785) putative amino acid transporter [Bacillus
Query Match 4.1%; Score 58.8; DB 1; Length 7983;
Best Local Similarity 46.9%; Pred. No. 6,6e-06;
Matches 183; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
QY 645 TCAAAAGTCTACTAGCAGCAACGCCATACGCTCCCTTAIAAAGCCCATATTGSCACCTT 705
DB 434 TGGAAAGGTATATTTAATGAAAACCTAGCTCTCTCTTAAGCAATGATAGAGTAAT 4400
QY 706 GAGCGTATCGGCAGAAATGTACTCAATCAATCTAGCGGCAGACAATCACGGCGGGA 765
DB 4401 AGGAGCTTGCCTCAAGGAGGAGGAGGATTAACTGGGAGCGGAGGTCCTCCACGGTGA 4460
QY 756 TATGATGTGGGGATATAGGACCAAGGAG:ATTACCTATCTGCCGGTACGTGGCCCTGG 825
DB 4461 CATGACACACACTTAAATAGCAGAGAGGAGTCTACTTCTCTGATTATTAGAGG 4520
QY 826 AGCGCGCGCTGTTATTTGGTGTATGCCCATGCTGTTCAGGSGTGATGTGTGAGATTTCGGGAC 585
DB 4521 ACCTTTGCTGCTCTAGGGAGCTTACACGCTTTAAAGGAGATGAGAGGATAGCGGTTTC 4580
QY 886 CGCACTAGAGTTTGCTCAATCACCAACATCAAAAGTCGATTTGATCAAGAAGTGGCAGCT 945
DB 4581 AGCGCTGAAGTGTGTGGAAAAGTCTTTTAGAAGTGGAAAGTCATAAAGGATTTGAATCT 4640
QY 946 TTCTTGGCCAGATGGAGAAATGCCGAAATATTATGAGTATTGSCAGTCGACGTCGCT 1005
DB 4641 CAAAATCTCTTGTGAAAACCTGCTGAAGTAAACAGCTACTATAAGCTTACGCGGAATCTTT 4700
QY 1006 GGAGGATGGCAGCGAAATGGCATATCGCGA 1035
DB 4701 AGCAAAAGCAGTGGAAATAGCGGTACACGA 4730
RESULT 8
LMFLCHR36_01/c
WPCOMMENT
Sequence split into 36 fragments LOCUS LMFLCHR36 Accession AL499624
Fragment Name Begin End
LMFLCHR36_01 1 110000
LMFLCHR36_02 100001 210000
LMFLCHR36_03 200001 310000
LMFLCHR36_04 300001 410000
LMFLCHR36_05 400001 510000
LMFLCHR36_06 500001 610000
LMFLCHR36_07 600001 710000
LMFLCHR36_08 700001 810000
LMFLCHR36_09 800001 910000
LMFLCHR36_10 900001 1010000
LMFLCHR36_11 1000001 1110000
LMFLCHR36_12 1100001 1210000
LMFLCHR36_13 1200001 1310000
LMFLCHR36_14 1300001 1410000
LMFLCHR36_15 1400001 1510000
LMFLCHR36_16 1500001 1610000
LMFLCHR36_17 1600001 1710000
LMFLCHR36_18 1700001 1810000
LMFLCHR36_19 1800001 1910000
LMFLCHR36_20 1900001 2010000
LMFLCHR36_21 2000001 2110000
LMFLCHR36_22 2100001 2210000
LMFLCHR36_23 2200001 2310000
LMFLCHR36_24 2300001 2410000
LMFLCHR36_25 2400001 2510000
LMFLCHR36_26 2500001 2610000
LMFLCHR36_27 2600001 2710000
LMFLCHR36_28 2700001 2810000
```



```

/translation="MGKITGSDLDVLLVLRKNEEDRNKALDFFYYEYVERELGEKVSYL
FDVKVLYEDERELPPKFLRDVARIK"
3144..3401
/gene="ST0005"
CDS
3144..3401
/gene="ST0005"
/Note="Swiss_Prot:P37413 percent identity: 52.030 in 25aa.
motif: phenyl group binding site (MAX box)"
/codon_start=1
/transl_table=11
/product="85aa long hypothetical protein"
/protein_id="BAB64947.1"
/db_xref="GI:15620951"
/translation="MSFSLVLRKWAYPFLSKVYNKESNLIPWVGSLSLKSLGK
VLRKLIYLLPHKGLAEIIVLKLILSSPFLSKSLIQCCMI"
complement(3247..3399)
/gene="ST0005"
CDS
complement(3247..3399)
/gene="ST0005"
/Note="P1R.S74018 percent identity: 58.696 in 46aa."
/codon_start=1
/transl_table=11
/product="50aa long conserved hypothetical protein"
/protein_id="BAB64948.1"
/db_xref="GI:15620952"
/translation="MSYNAEEFLPRAXDYLKASLILFQQGILYFASALDSEVSARLEI
NFLNIG"
3744..3483
/gene="ST0002"
CDS
3744..3483
/gene="ST0002"
/Note="Swiss_Prot:P55577 percent identity: 37.004 in
227aa."
/codon_start=1
/transl_table=11
/product="579aa long hypothetical prolyl endopeptidase"
/protein_id="BAB64949.1"
/db_xref="GI:15620953"
/translation="MDEYEYLN.SDPRTISFIEKREKTESKLGKRAKAMELYPLLELM
DREPYLSMFAFDEENPAILLYGEKSQLLNKNTIYIPPEGYVASEIKVYNKSEIGV
STYERKSDKIITLLSPGKIRLEGMVESPFYKGECLYIKSYRSPDPGGYPAD
RIFKDEDAIVYGDMPKPGFVTIKVDFDFTLVROKGRWYGLYVGEVDSLRKYVDEGE
VIDIDFOGGEVIYQNNVAVLGNKVVEVDYPVIGVSHIGDKTAVEVKEYRPLIF
YDKGKKIGEEVHNDITWDKGGHILFLVPSFNKFKRVVAKSDEGRVILMOGNYI
VTVKLYVKGVLHMGFLVSRANPKGVIVYGGFRPLPLPSFVSVMRVLINSEYIS
LITNIRGYENGEEHMKAGMLLNKKNFDFAEFLRLVKIMCGKAVAMGGNGGLVG
ATINYEPLIDCAVLIGHVLDMLRYDKLYGVKYVVEYXGDPNDPKYTFLLSYSPHN
LKKGLPTFTVTGINDRVIHPHALKLYVAKSKSLGNDVNLVFNUSGHSIADPESKAP
ESYVVSFTDECLR"
complement(5570..5819)
/gene="ST0003"
CDS
complement(5570..5819)
/Note="Swiss_Prot:Q40469 percent identity: 42.846 in 67aa.
motif: ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="949aa long hypothetical protein"
/protein_id="BAB64950.1"
/db_xref="GI:15620954"
/translation="MICSSSTLSLDAVFKFSDSGVPTTNNVKFPLVLTLSGN
VRLCDLNVYOHGLFALNIMGKNLVVNAFTGGKTEIWSVVALEMQLKGFENLAIY
PTKALAGQOIERVAVYIENGESVSKVQKTKIVEYVGVKIGLYSYNYSIQSVKR
AKTLINPEVKNALFOHKKISDFLKKVRI:VWDFPFYSGSKSTVILITKIDKP
GIKPOIVINATSLDOPEVVKPFDFEIIIGKSGFERANDYIYVIGKDELNIAISKAJY
SFDDIVKVFKEFYASDKELIYQLFQDNAOILFYLQKKGDELTIFSRSSFEAN
NLVGLKLGKVNVDGTAPVAIVHSGIDKYVROOVENDMSGKLKVVTYKTKGGIDVS
NVRVHVHGIPODSVEFGREGKGRMASIKKTESVIFPISLSQAVLLEDSTSLKFN
LSLGSFSLILLPNPFKLIDV:RGILNDREFLKSGLTICIPQISFTEQHLKTPKL
LFDGKSNVVKLNLRDVERKQGCIDPMLNAIWRNSVINGRHHVETISDFSSLM
CNINISNVKNAIYAYEDICFSWKQHPDIENDRGKVMKSLDVLFDSDGFKQ
VHEIARKFWYIFESRKKIGOEKVEKIELEVTYPSNIAKYDFTLYVYASELEPDDVVKY
DKGNVFIALLRLNVDGLGLNFGVSGILK1WESEPTALLKRLMKKVKVKGIE
```

```

LCKKLLDDVRNAPQFRRLVLVLYDYTFNNDFNEVKAIARFVHYLCNTPICL
QVIGTMIPKSPILRSIIIDSPGKYAVSSPTGEISIFDNEEBAVKAALNIEFTDA
IVVPYGSICKYKQNEVYIDKEINRLGGPIIPSKFRELVLNCDLSLKEENADN
SITRGEVHLTKIPRKAETIRILNMLWKAYLTK"
9382..9549
/gene="ST0007"
CDS
9382..9549
/gene="ST0007"
/Note="OWL:AF072619 percent identity: 27.500 in 40aa."
/codon_start=1
/transl_table=11
/product="55aa long hypothetical cytochrome b"
/protein_id="BAB64951.1"
/db_xref="GI:15620955"
/translation="MKACTPSLKHENFELIYMI.TPISLKRSSRLDITISPSRQEQSL
CILNIYELFYI"
complement(9819..10127)
/gene="ST0004"
CDS
complement(9819..10127)
Query Match: 3.94; Score 55.8; DB 1; Length 263050;
Best Local Similarity 52.94; Pred. No. 0.0001;
Matches 120; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
Oy 710 STATCGGCAGAAA.TGAM:ICANTCAATTCACCTGACGCGACAGACAAICACGGGGGAATAIG 769
25 172729 SCNTTACNTATAGGAATAATTGAGTACCAIACCTCTCTAGAGAAACGGAGGAATAIG 172670
Qy 770 GATGTCGGGATATAGGACGAGGAGTATTACCTATCTGCGGGTACGCGCCCTGGAGGC 829
2b 172669 GATATTAAGCACITTAACCTATAGGCACATAATATACCTTCCAGTATTGTAAGGGAGCA 172610
Qy 830 CGCCTGTTATTGGTGATGCCCATGCTGTGTACAGGTGATGTGAGATTTGCGGACCGCA 889
2b 172609 TTGCTTCCATAGGTGACACTCACTTACCCCAAGTGATGCTGGAAGTTGCGGAACGCA 172550
Qy 890 GTAGATTGCTCAATCACCACCATCAAACTGATGTTGATCAAGAA 936
2b 172549 ATAGAGCCCCCTGGAGTTACANTGAGTTAGTTAATAAAAA 172503
RESULT 10
SME591789/c 294800 bp DNA linear BCT 05-JUL-2002
LOCUS Sinorhizobium meliloti 1021 complete chromosome; segment 8/12.
DEFINITION AL591789 AL591688
ACCESSION AL591789.1 GI:15074950
VERSION AL591789.1 GI:15074950
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti.
Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae: Sinorhizobium.
REFERENCE 1 (bases 1 to 294800)
AUTHORS Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Gedrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetecole,D., Puchler,A., Purnelle,B., Ransperger,U.,
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
ANALYSIS of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
JOURNAL 21395507
MEDLINE 11481430
REFERENCE 2 (bases 1 to 294800)
AUTHORS Gouzy,J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELLILLO
EJ Consortium
COMMENT
MELLILLO EJ Consortium:
Laboratoire de Biologie Moléculaire des Relations
Plantes-Microorganismes, UMR2115-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genétique et Développement UMR6061-CNRS,
Faculté de Médecine, 2 avenue du pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
```

Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,  
D-33615 Bielefeld, Germany. Unite de Biochimie physiologique,  
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,  
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte  
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,  
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr  
<http://sequence.toulouse.inra.fr/mellilot1.htm>.

## FEATURES

```

source
    location/Qualifiers
    1..294800
    /organism="Sinorhizobium meliloti"
    /strain="1021"
    /db_xref="taxon:382"
    complement(147..1382)
    /gene="SMC04259"
    complement(147..1382)
    /gene="SMC04259"
    /function="cell processes; transport of small molecules"
    /note="Product confidence : putative
    Gene name confidence : hypothetical
    predicted by Codon_usage
    predicted by Homology
    predicted by Framed"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="PUTATIVE PERIPLASMIC BINDING ABC TRANSPORTER
    SIGNAL PEPTIDE PROTEIN"
    /protein_id="CAC46508.1"
    /db_xref="GI:15074951"
    /db_xref="SPTREMBL:Q92P64"
    /translation="MNRFLAAALGATAALPFGAASATDLEVIHWTSGEAAVAFI
    AKAFATGNKWDGALAGSGGTARPIKISRIITGDPMMAATOFNHRGROAEELVQAGLMK
    DLDTIAENKKEIVKPSLLDSCTIEGKIYCAPVNLISKQMWLSNAPKQACQVEYP
    KNWDEVAARALEKAGSIVPLAVGGOPHOANGAFDVLVATAGAKENFEKVFPAQKDEEV
    RAGPTKVFRAADCAQSMKACTNVDMNOANGATVITKAGGONGWAGGEPOLARQ
    KAGVDTICPLGLVNEVISTGDAPFPLLEDEEKSKAQEVLASILLKPEIQVAPNLK
    KGSLEVRGVDAIDAAANDCKMKGDLILAKGNVIQGTDLLSADSOKOKEDLFSEPFANI
    SMTPEDAKRRFDIIIAAAD"
    1858..2928
    /gene="SMC04260"
    /1958..2928
    /gene="SMC04260"
    /function="miscellaneous; not classified regulator"
    /note="Product confidence : putative
    Gene name confidence : hypothetical
    predicted by Codon_usage
    predicted by Homology
    predicted by Framed"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
    /protein_id="CAC46509.1"
    /db_xref="GI:15074952"
    /db_xref="SPTREMBL:Q92P63"
    /translation="MTGMDGKTKRATTAAPVEGGRPLKTIATFTGLGVTVSKALK
    DAPDLAGTKERVLKAIQGNQNRAGVRLRTGKTVNIVSIWLTLESEIGITSPVMI
    GITELAGTVHVVTVYSSAKDPLGPIRYILDGAADGVILISKTEPNDRVLTITET
    RLPEATHGTREGLIHPYEDNEFEAYEAVRKLVDGRRLRLERPPNLTTHSHMK
    TGFERGVDRGNAVSPHQVNDISILVAIRDAPEKLRSPADPCVVSQSGSGAJALI
    AGIEAAGKKVDEADSKVPSDFLRLWRPEVNTMYEDIRIAGRELAKAVTGTIEGRP
    PTTIOSLOPEQPMARSAYS"
    2954..3023
    /note="Sm-4 OR SMC04506
    REPEAT SM-4
    Predicted by Homology"
    /evidence=not_experimental
    complement(3036..3302)
    /gene="SMC04261"
    complement(3036..3302)
    /gene="SMC04261"
    /function="miscellaneous; unknown"
    /note="Product confidence : hypothetical

```

```

Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL PROTEIN"
/protein_id="CAC46510.1"
/db_xref="GI:15074953"
/db_xref="SPTREMBL:Q92P62"
/translation="MKFTVIFGGKAAEVTAAGSVTIRDEAERRIKAAAGVEQWRVRSIA
TGAPTPGIQYLKMOIDFVAGKLEQLNPISDFTDDKYWPAAGI"
complement(3407..4837)
/gene="gnd OR SMC04262"
complement(3407..4837)
/gene="gnd OR SMC04262"
/EC_number="1.1.1.44"
/function="small molecule metabolism; energy metabolism,
carbon: oxidative branch, pentose pathway"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE 6-PHOSPHOGLUCONATE DEHYDROGENASE
(OPCARBOXYLATING) PROTEIN"
/protein_id="CAC46511.1"
/db_xref="GI:15074954"
/db_xref="SPTREMBL:Q92P61"
/translation="HQAEGIGLIGLVMSGNLALNAEKGNITAVENRTVDATRKFYA
EGALQKQIVPCETIEEFVAAIRPPRIILMIKAGDPVQDOQNEALKPLHAKGDMIDA
GNANFDIWRREDALDKSLGTLFIMCVSGSGESGAGHPSPIMVGGTEESTRRKCVLTS
IAKYDSDCVLANLGENGAGHFVKTIHNGIEYADMOMIAEIIYIURDLKMTAQETGE
VFGANWKRGLNSYLIETEKVLAADPLTGKPIVMDILDKAGCKGTGKMSVIAQNMG
VPATAEAAVAARSISSAKEEREAAKVLGPVGEIKVADRDGTIKDENALLAAKI
GAYAQGFAYAAASKEFGNLMPTAKIWRAGCIIRSFOLDEITAITAFKAPDAANLI
VTPFAAMVKESDGLARRVSTAVLGELPVPALASALGYEDSYRRGRGTANVIOQRD
FTGAHGFDVRDGDADSHHGPWGSLNA"
5273..6691
/gene="SMC04263"
5273..6691
/gene="SMC04263"
/function="cell processes; transport of small molecules;
amino acids, amines, peptides"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE AMINO ACID CARRIER TRANSMEMBRANE
PROTEIN"
/protein_id="CAC46512.1"
/db_xref="GI:15074955"
/db_xref="SPTREMBL:Q92P60"
/translation="MDTIIGLNTIFWGVVLYGLLAVGLYFTVRLGFFQIVHFSEMF
RVLSRSGSKDAAGISPPQALMVSLASRGVTGNLAVVALYLGPGATFMMWIVAFVG
MATAYAESALAOIYKIRNEDGOVYRGGPAPFYIARGNAPMAATIFSACLILSFGIVFNA
VOANSTADAVOGAFGVPKLVAGVAVALLSGVIFGGIROIARVAEITVVPMAAAYLI
AVVLIANASAVPGVLTWLTISSAFGEQAEAGIIGGIAAMLNGVKRGLFSNAGMGS
APNIRAVATVPVHPHSSQGFVQSLGFTIDILICTATSVMLLSGLFPGSVGTGTOL
TOAMSTHIGSAGTYFIATIAFFFAFTSIIGNYSYAENALTYLGGNRLGLMIMRCAT
LANVVMGAYESITIVFDAADASMLMATINLVAILLLSGTIAKLTNDYFEQRKAGAVP
VFHAADYPELQKIDGINSRD"
6773..7663
/gene="SMC04264"
6773..7663
/gene="SMC04264"

```







## COMMENT

Road, Exeter EX4 4QG, United Kingdom

On Dec 16, 1999 this sequence version replaced gi:6224591.

Notes:

Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.

(URL: <http://www.sanger.ac.uk/projects/s.pombe/>)

During 1995 to 1996 about 66% of *S. pombe* chromosome 1 was sequenced by the Sanger Centre. The sequencing of the *S. pombe* genome is now being continued with funding from the European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.

Protein coding regions (CDS) have been predicted with the help of computer analysis using the GeneFinder program in PomBase (an ACEDB database), with additional predictions for the branch-acceptor sites supplied by the program SplicePlot. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.

CDS are numbered using the following system eq SPAC25H2.CDS.SP (S. pombe), B (chromosome 2), c25h2 (cosmid name), .01 (first CDS), C (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Location/Qualifiers

Source

1. .35346

/organism="Schizosaccharomyces pombe"

/strain="972h"

/db\_xref="taxon:4896"

/chromosome="I"

/map="IR"

/clone="cosmid c869"

/complement(1..1610)

/gene="SPAC869.11"

/partial

/complement(1..1610)

/notes="SPAC869.11"

/note="SPAC869.11, len:536, SIMILARITY:Schizosaccharomyces pombe, ISP5\_SCHPO, sexual differentiation process putative amino-acid permease isp5., (543 aa), fasta scores: opt: 1812, E():0, (53.9% identity in 492 aa)"

/codon\_start=1

/label="SPAC869.11"

/product="amino-acid permease"

/protein\_id="CAB60021.1"

/db\_xref="GI:6224602"

/translation="MEPEVVSFSDKTKILNESKSLKDKVPSLEKSYTFGLVDD VPKGNFVRFDDPKPAKATGEDTALKRSLKSRMQMISIGAGTGLYVGRGSS LADGNFASVIVNLSLIGIMHFTIVAGLGMAYVPVAGFNTATRFIDPANGFAVSS NYFINVFPELLETTCAITFRYWDINSAAW-STPLVVIITVIVFGVAYGVGFIL STVKVATGFGILAIINCGVPIDHRYIGSIKKHPRGPGSPCTIAAFS SVCTVGLAAEVDMPKALPHAVKQVPRIAIEVVSLLIGLLSPDDNLNMG SIVSPVPLAEANIKGLPSFVNAV-IISVVSVINSTYTAGRTLHGMAKLRQKPSF FYTRDRLPLAMIVLLFGFAYINEADKGNVDZVFNKLALSGESNFTVGS IGLHIFRLAKKQGSKELGFSVPMGIWGSICGLGFNILLCKAQFVVSLEPAGGX PNANDFGYLAAPVTAFVFGYKIYDRS"

1. .1141

/note="Overlap with: c922 S. pombe chromosome 1"

/complement(3..1394)

/gene="SPAC869.11"

/note="Match to PF00324 aa\_permeases, Amino acid permease score 820.26"

3705. .5363

/gene="SPAC869.10c"

3705. .5363

/gene="SPAC869.10c"

/note="SPAC869.10c, len:552, SIMILARITY:Emicella nidulans, PUTX-EMENT, proline-specific permease, (550 aa), fasta scores: opt: 1968, E():0, (53.3% identity in 520

aa)"

/codon\_start=1

/label="SPAC869.10c"

/product="amino-acid permease"

/protein\_id="CAB60020.1"

/db\_xref="GI:6224501"

/translation="MDKKALQYBESKRAREITDIELVSGGIDVEKKYGETKRAIKS RHVQLIAGCGICIGTCLFVSGSSALSESQSPASILSYVIMSVIIVNALGEMCTYLP LSGASPTIIBERYDASIDAFAGANNYAYVFLVASEVTAASIVIEWTYATVPAGWLI AILFLVAVLNSFEVFKMGFTFWFAIKVIAIVGLIILGVVIFFGTTPKDRGLGFRY WKHLIFAVLNSFEVFKMGFTFWFAIKVIAIVGLIILGVVIFFGTTPKDRGLGFRY SRTVIR:FFYIFGSLTIGTIVTSKDPRLNAISSGASGAASPEVIGIQAEE.PVLN HIINAVILTSACSSGNSHLPAGSISVSLAKEHOAKPIFYKCNRWGVVSVAVTVLF ACLAFNASASAAVFWFNCNLSTLSGLAMICVLYAYLOPKAMILNINLWETRPYKT PWSFYATLITLLALITLNGFVFGHTTAGNTAAATYITLIFLVLVVAHKLWSR NNSFGKRIEEDTIGTVAEAELEOMYPAPVPRNIIEKINFWIA"

3816. .5228

/gene="SPAC869.10c"

/note="Match to PF00324 aa\_permeases, Amino acid permease score 18.22"

complement(6788..7138)

/note="TIF1 ITR"

/complement(7524..7874)

/gene="SPAC869.09"

/complement(7524..7874)

/gene="SPAC869.09"

/note="SPAC869.09, len:116, SIMILARITY:Schizosaccharomyces pombe, YAAB\_SCHPO, hypothetical 15.4 kd protein c24g7.11c in chromosome 1., (143 aa), fasta scores: opt: 192, E():2.5e-06, (39.6% identity in 139 aa)"

/codon\_start=1

/label="SPAC869.09"

/product="hypothetical protein"

/protein\_id="CAB60019.1"

/db\_xref="GI:6224600"

/translation="WPDZAHIIAGHKAALSNPHVSEAKRARKYLKEHGESHSHTTG TTRGOKADADDAGELRFRFGCTKQGFENESGANNVGRGCKYKAAMHNPAVSKGRR HAKSLREVDDESK"

complement(8736..9448)

/gene="SPAC869.08"

/complement(8736..9448)

/gene="SPAC869.08"

/note="SPAC869.08, len:230, SIMILARITY:Triticum aestivum, P1MT\_WHEAT, protein-1-isoaspartate O-methyltransferase, (230 aa), fasta scores: opt: 602, E():2.2e-32, (48.5% identity in 200 aa)"

/codon\_start=1

/label="SPAC869.08"

/product="putative protein-1-isoaspartate O-methyltransferase"

/protein\_id="CAB60018.1"

/db\_xref="GI:6224599"

/translation="NWSFNLSNNAALVOHLVESKFLNORAIKAMNATSRSFYCLPS PYMDSQSIGTGVITISAPHMATLORLEPVLQPCSCALDIDGSGYLVAAMARVAP NCTVGEIIPOLVETSKKLLKDINHDELMYKERRLOINVDGDMGTSEDEKFD ATHVCASASE.PQKLVDDLKSPGKILIPIGTVSONIYLIEKNEQGIKSKRTLFPPVRY FLIDSPDSSSDY"

complement(8783..9427)

/gene="SPAC869.08"

/note="Match to PF01135 PCMT, protein-1-isoaspartate(D-aspartate) O-methyltransferase (PCMT) score 263.14"

10265. .11575

/gene="SPAC869.07c"

10265. .11575

/gene="SPAC869.07c"

/note="SPAC869.07c, len:436, SIMILARITY:Mortierella vinacea, G02402, alpha-galactosidase, (417 aa), fasta scores: opt: 1443, E():0, (52.3% identity in 407 aa)"

/codon\_start=1

/label="SPAC869.07c"

/product="putative alpha-galactosidase"

/protein\_id="CAB60017.1"

misc\_feature

gene

CDS

```
/db_xref="GI:5224599"
/translacion="MISISELNOPELVPLFESDVGSYNGLQKXQWNSKNKYA
CDIDESLIUNNAKAKKEGLICIGYHYIMDCSKSRNATIGLILANWKPENGIG
SMARKLHMGKPFMYKMSDALNKGRTFYISCOMGEFVNMNGNTIANSKRISSEFJ
FSRKLIVAPCPCTIECFALOGHCSVMNIISKASFLSSKAGNSGNWJESIEVNGGM
SFEYKTHFTWAILKSPILGNEVSSNSPMUKLISNKKELISINQDIGNFAALINW
KKYGEVTELFSGSLNSNDWVAVLNAASEPLKNGIHLSDIFVIALGNARHDWATDL
WNNVKKLVSDPIKANVASHGVQVVRFPQYKYKNTNCKFEFSNKI"
13058..13869
/gene="SPAC869.06c"
13058..13669
/gene="SPAC869.06c"
/len="203"
/codon_start=1
/label="SPAC869.06c"
/product="hypothetical protein"
/protein_id="CAB60016.1"
/db_xref="GI:6224597"
/translacion="MLQYSKKKVSINFPVRLLSYKMPISDAFKOHRKLOSDYQNI
KSANDYDTATWONOFVWELARHSVGEIIVVVKFYLGESKEMAEKDRHEHLVK
EMLYKFOSKMANOSNFIPALDEMSLOKHIDEEEOHDIPLEKHLSEESLHAKSF
ERIKKFTPHSRPSAPNPFPETVAG:FAAP:DKLROMKRW?"
16383..18905
/gene="SPAC869.05c"
16383..18905
/gene="SPAC869.05c"
/len="840"
/note="SPAC869.05c, len:840.
SIMILARITY:Schizosaccharomyces pombe, SULH_SCHPO, probable
sulfate permease spbc3h7.02.. (877 aa), fasta scores: opt:
3127, E(1):0.752.8% identity in 827 aa"
/codon_start=1
/label="SPAC869.05c"

Query Match 3.0%; Score 42.8; Db 8; Length 35345;
Best Local Similarity 55.3%; Pred. No. C.63;
Matches 83; Conservative 0; Mismatches 67; Indels 0; Gaps 0
```

```
QY 729 CAATCAATTCATGACGCGAGACATCAGCGGGGGAATGATGATGCGGATAGGAC 753
DE 22212 CAGGACATATCCGGGACACCGGAAATGGASGAATTCGATATTAAATTTATCAA 22153
QY 789 CAGGAGATATTACCTTCGCGGTACGTCGCCCTGGAGCGCGCGCTTTATGCTGATG 848
DB 22152 GAGGTCGGAAGATTTTCTTCGCGTCCAGCTTCCTGCTGAATGTCATTCGCGATC 22034
QY 849 CCCATCTTGTACGGGTGATGTCGATTT 878
DB 22092 TTCAITTTTCTCAAGGTGATGCGGAATTT 22063

RESULT 13
CNSPAX03
LOCUS 307120 bp DNA linear BCL 0%-SLP-1999
DEFINITION Pyrococcus abyssi complete genome; segment 3/6.
ACCESSION AJ248285 AL096836
VERSION AJ248285.1 GI:545807
KEYWORDS complete genome.
SOURCE Pyrococcus abyssi.
ORGANISM Pyrococcus abyssi
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
1 (bases 1 to 307120)
Heilig,R.
Pyrococcus abyssi genome sequence: insights into archaeal
chromosome structure and evolution
Unpublished
2 (bases 1 to 307120)
Genoscope.
Direct Submission
Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
```

```
FEATURES
Source
..307120
/organism="Pyrococcus abyssi"
/strain="Orsay"
/db_xref="taxon:29292"
98..1144
/note="PAB0437"
/codon_start=1
/transl_table=11
/product="aminopeptidase from family M42"
/protein_id="CAB49557.1"
/db_xref="GI:5458068"
/translacion="MVNYELKKVVEAPGVSGYFELSDVWIEIKDYDEVKVKDL
GNVIAHKGEGPKVMAAHMDQIGLMTNIEKNGFLRVAPIGGVDEKTLIAQFKVMI
DKDFTYGVGASVPHIQPEERKKAPEMDQIFIDIGAESKEAEEMGVRIQVITWD
GRLEGLKHREVSIAEDDRIAVYTMLEVAROLEDTKADYVFNATVOEENGLRGATSA
FAEPDYGFAIDVTIADVEGTPEHKQVTHLGKGTAKIMDRSVICHPTIVRWLEELA
KKYEIPVLEILLGGTIDAGAIHLTRAGVPTGALSVPARYIHSNAEVDEROVDATVK
LMVKALENIHEIKI"
1188..1535
/note="PAB0438"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49558.1"
/db_xref="GI:5458069"
/translacion="MKVYKENVKVKLVTRTKRGAILHMIELKPGHFLEQNPLKPSKY
GEAYRTKGNFPPEFYFMEIKDNKYTRILAGAFLEKEIDEFLTLLAKTEDFKKLEH
VLEIEIEEGEE"
1570..1656
/product="tRNA-Leu"
/note="codon recognized: AAA"
/anticodon=(pos:601906..601908,aa:Leu)
1674..2003
/note="PAB0439"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49559.1"
/db_xref="GI:5458070"
/translacion="MMPMNPOLKJLMKOLDKMDGKVKVILKENKEIVIKPVVT
VIRANGKMQIAGTTEERVVLKISEEDIKLWMEQAGVDYEITAKKALEAGGDLAA
ILRLTDQ"
complement(1983..3347)
/note="PAB1940"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49560.1"
/db_xref="GI:5458071"
/translacion="MSVYDEYLDNKGWIRKALEIVRFKKEKLEVIKRALQEKDPL
PEDLSSELDVLIGIVENKPEPINEYAAEGVATLGAKIIMARKIFSWIRGILQKSI
VKGTLYIKKIAMKKELSTNEIRELMPSLIRIILDDKDESVVLOTVETINAIINPMDL
KKYSALRVREIVTRTKNEKIMEEAFNLALDVNDPPVRSVSTALKILEVIDRREF
LSRESLEKASMLSSSIGRTKEKIDELIQGREKKVTEVSKDVSVEIKEMFER
EOHMAVLAKSDYKVLQVIRIFLSEDLTKDALWVSNATYSIKIDAKLTIPOL
PEFLHYNPWVSTSAKVALALEYSIMEPEISKALIESNDEKLVLAGTELADA
LLSRKIKGTFLREVMKRLIRKDRLPREALDPIKKYESYIELEPEIKRQ-SIKLVNLI
VGEF"
complement(3344..4747)
/note="PAB1939"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB49561.1"
/db_xref="GI:5458072"
/translacion="MVEGFSIVNVEQIRDDLKKWKVREVLESIDIAKKVEILKKLLT
DNKTVGNTLFIETIRIKGKLDSEQLDITIRLSKSSNDKIALNSLKLRTIL
DNVQLSNGYDKVINTLSGILTSRRGILREYAAEELGILGAKITSFKRIVEFLVSLI
KEGDRVKGAALSALTIASRSDNVEVIDITREIRIKTEODAEKKATISLSLEI
IFSKRKOKLTPKTMSEIKEEVPEIDKSTVEPYEENVNDIEKLFEMEKHEVATLAKENE
GILKKVDEMLSSKEYIRRADALVWARIKFLDPNMAVYVKNLAGLVKDRNPWRNT
AIKATAEAYLVPEVRMDVIPLLDALLRSNNKNDVELALNIIRKIELSTSHDEELFRAT
```



```

JOURNAL Nature 404 (6777): 502-506 (2000)
MEDLINE 20222556
PUBMED 10761919
REFERENCE 2 (bases 1 to 349061)
AUTHORS Parkhill,J.
JOURNAL Direct Submission
TITLE Submitted (30-MAR-2000) Submitted on behalf of the Wellcome
sequencing team, Sanger Centre, Wellcome Trust Genome Camp.s,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
URL: http://www.sanger.ac.uk/Projects/N\_meningitidis/.
FEATURES
source
1..349061
/organism="Neisseria meningitidis 22491"
/strain="22491"
/db_xref="taxon:122587"
/note="serogroup: A"
complement(24..206)
/note="ATR repeat: hmufs hit to HMM ATR (1 - 183). score:
310.39"
/label=ATR
209..212
220..681
/gene="NMA0368"
220..681
/gene="NMA0368"
/note="NMA0368, probable integral membrane protein, len:
133 aa; contains four probable transmembrane domains"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83669.1"
/db_xref="GI:7379121"
/db_xref="SPTREMBL:O9JWH7"
/translation="MQEQRKSSFP:YMLLYSVLWIASLSNVAFYLNHGSMEGLIV
LILGSFASLDIRYCAVANYWLAITVLEALRKVVPVHAFWGLA:VAFSVKVVY
DRAGTSDIVRYGAGFYLMWYAFVAVASITGAGKPKKAASADGTRNDY"
671..1492
/gene="hemK"
671..1492
/gene="hemK"
/note="hemK"
273..277
/note="hemK protein, len: 273 aa; similar
to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa). fasta
scores: E(): 0, 42.3% identity in 279 aa overlap. Contains
P300092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/product="hemK protein"
/protein_id="CAB83670.1"
/db_xref="GI:7379122"
/db_xref="SPTREMBL:O9JWH6"
/translation="HTFDKWLGLSKIPKNEAPMLIQYVSEXTRVQLTRGGEMTDFV
RQADRLAQRRINGEPFVAYIIGAREFYGRRTVNSVLIIPETEH:WEAVCARIPEN
GRWDLGTSGGAAVAVTARDPDAVRASDISPPALERTAKNAADLGARVFAFGSWE
DTDMPSEGGKRD:IVSNPPYIENGDKHLISQGLRFEFGALIDFSDLSCTITLQAGAE
DRLAEGFLLH:HGFDQGAAVKGVLAENGFSQVETIPDLASLDRTVLSKYNKHLK"
831..840
/gene="hemK"
/misc_feature
/note="Core DNA uptake sequence: ggcgtctgaa"
/label=DUS
1162..1171
/gene="hemK"
/note="Core DNA uptake sequence: ggcgtctgaa"
/label=DUS
1187..1207
/gene="hemK"
/note="PS00092 N-6 Adenine-specific DNA methylases
signature"
complement(1279..1288)
/note="Core DNA uptake sequence: ggcgtctgaa"
/label=DUS
1565..1568
RBS
1578..2966
gene
/sequence="NMA0370"
1578..2966
CDS
/note="NMA0370, probable integral membrane protein, len:
462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
P44640 hypothetical protein H10325 (450 aa), fasta scores:
E(): 0, 49.8% identity in 464 aa overlap"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB83671.1"
/db_xref="GI:7379123"
/db_xref="SPTREMBL:O9JWH5"
/translation="KNVAVAVVIMLVLSLSRVHVVLSITIGAFYGVGAGMPLONIA
DRAGVSOAGIIPVFNKNGLEGGAKTALSYMLGAFAMATHSGLPQQLAGAVVRKLNK
GMPDSVRSGEGVWLLLSIILVMGMSON:PIPIHAPITMIVPLPPLLVFNRLKIDR
RLIACVITFGLVNTYMFYGFGLFNLILLGNHSAPOLDVKNINVMAMATPAL
GLIAGLLAFVHYRKPRLYGSNNADTAGNADANRQPSAYRSLAAVAIAVCFAIQI
MYEDSLVGLMGLGFAVGMKLVNIRKANDVFEGIKMMAMVGFITLAAQGVFAVMA
TGHIOPLVFSSMAIFGNSKMAALMLVVLVTMGISGSRSTFLIILAIYVPLCVGL
GSPSLATVAVGTAGALGAGSPASDSTLGTPTMGLNADQHDHRSVIPFIHINIP
LLIAGWIRAWVL"
2967..3191
gene
2967..3191
CDS
/gene="slyX"
2967..3191
/gene="slyX"
/note="NMA0371, slyX, SLX protein homolog, len: 74 aa;
similar to SLX_HAEIN P44759 SLX protein homolog (73 aa),
fasta scores: E(): 0.0017, 33.8% identity in 74 aa
overlap, and SLX_ECOLI P30857 SLX protein (72 aa), fasta
scores: E(): 0.24, 32.4% identity in 68 aa overlap"
/codon_start=1
/transl_table=11
/product="SLX protein homolog"
/protein_id="CAB83672.1"
/db_xref="GI:7379124"
/db_xref="SPTREMBL:O9JWH4"
/translation="MDAVQEFHRIITEIQSALOEDVIAGLNMVAVELROTLDLQQA
QLRLLYQKMOORNPDAQEYSLRDETPHY"
complement(3271..3423)
/gene="NMA0372"
complement(3271..3423)
/gene="NMA0372"
/note="NMA0372, unknown, questionable CDS, len: 50 aa"
/codon_start=1
/transl_table=11
/product="very hypothetical protein NMA0372"
/protein_id="CAB83673.1"
/db_xref="GI:7379125"
/db_xref="SPTREMBL:O9JWH3"
/translation="MEGRSLPLGSDGFF:LAVVEYVPVGFICLOGKTAIFRSQGRPHRI
KGI:SD"
3562..3571
misc_feature
/note="Core DNA uptake sequence: ggcgtctgaa"
/label=DUS
complement(3586..4356)
/gene="thif"
complement(3586..4356)
/gene="thif"
/note="NMA0373, thif, probable Thif protein, len: 256 aa;
similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
fasta scores: E(): 0, 43.1% identity in 246 aa overlap,
and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
protein. (249 aa), fasta scores: E(): 0, 43.9% identity in
244 aa overlap (note that N.m. does not have orthologs of
any other molybdopterin biosynthesis proteins). Contains
pfam match to entry pf00899 Thif_family, Thif family"
/codon_start=1
/transl_table=11
/product="Thif protein"
/protein_id="CAB83674.1"

```

```

/db_xref="GI:7379126"
/db_xref="SPTREMBL:Q9JWH2"
/translation="MTTTEHNDADFLLRYSRHILLDEIGEQQKLSAAH:IVVGGG
GLCAAAALPYLAASGIGTITLADSDTVLHNLQROVAFDGVGKLTFLADRIKRIIN
HVVDVITNEKLDGGRGLGVQADIVLDDCDNYATPOAVNRACVCAKTP:VSGAAVR
PQGLAVR:PDLPSQPCACLEDGKSASDNC:SLPQVPSLVGITIGSTQAA:ALKLL:
DAGPESHGRLAVYRALEGSGWQTFD:PRNPSP:VCGAPR"
misc_feature
3787..3796
/note="Core DNA uptake sequence: gccctctgaa"
/misc_feature
/label=DUS
complement(3847..4257)
/gene="thiP"
/note="Pfam match to entry PF00694, thiP_family, ThiP
family, score 136.60, Evalue 4.1e-52"
/misc_feature
complement(4416..4427)
/note="Core DNA uptake sequence: gccctctgaa"
/label=DUS
4471..7224
/gene="ppc"
4471..7224
/gene="ppc"
/gene="ppc"
/PC_number="4.1.1.31"
/note="NKA0374, ppc, phosphoenolpyruvate carboxylase, Gen:
917 aa; similar to many e.g. CAPP_PHOA Q32483
phosphoenolpyruvate carboxylase (PC 4.1.1.31) (936 aa),
fasta scores: E(1): 0, 43.3% identity in 928 aa overlap.
Contains 2x Pfam match to entry PF00411, PEPcase,
Phosphoenolpyruvate carboxylase, PS00017, ATP/GTP-binding
site motif A (P-loop), and PS00393, Phosphoenolpyruvate
carboxylase active site 2"
/codon_start=1
/trans_table=1
/product="phosphoenolpyruvate carboxylase"
/protein_id="CAB83675.1"
/db_xref="GI:7379127"

```

[illegible]

RESULT 15					
AX041919				DNA	linear
LOCUS					
DEFINITION	AX041919		Sequence 814 from Patent WO065062.		
ACCESSION	AX041919				
VERSION	AX041919.1		GI:11340678		
KEYWORDS					
SOURCE	Pyrrococcus abyssi.				
ORGANISM	Pyrrococcus abyssi				
	Archaea; Euryarchaeota; Thermococci; Thermococcales;				
	Thermococcaceae; Pyrococcus.				
REFERENCE	1 (bases 1 to 343980)				
AUTHORS	Forterre,P., Thierry,J.C., Prieur,D., Dietrich,J., Lecampre,C., Querellou,J., Weissenbach,J., Saurin,W., Hettich,S., Fiamant,E., Raffin,J.P., Henneke,G., Guequen,Y. and Rolland,J.F.				
TITLE	Genome sequence and polypeptides of pyrococcus _i(abissy), fragment and uses thereof				
JOURNAL	Patent: WO 0065062-A B14 02-NOV-2000; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IPREMER INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)				

```

Location/Qualifiers
1..34980
/organism="Pyrococcus abyssi"
/db_xref="taxon:2929"
/note="Original length of seq 1: 1765118 (for 812
sequences) New sequences : seq 001, from 0.000.001 to
0.349.980, length: 349.980 seq 813, from 0.300.001 to
0.649.980, length: 349.980 seq 814, from 0.600.001 to
0.949.980, length: 349.980 seq 815, from 0.900.001 to
1.249.980, length: 349.980 seq 816, from 1.200.001 to
1.549.980, length: 349.980 seq 817, from 1.500.001 to
1.765.118, length: 265.118"

BASE COUNT      94050 a 78692 c 78319 g 98879 t
ORIGIN

Query Match          2.9%; Score 41.2; DB 6; Length 34980;
Best Local Similarity 53.8%; Pred. No. 2.9;
Matches 85; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY    743   ACSCCACAAATCAGCGGCGAATATGGATGCGGGAATAGGACCAGGAGTATTACC 802
      ||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB    282490 ACCCCGGAACCAACGCCGGAAACATGACACAACCTTAATCAGAAGGGAACCCACCAT 282449
      ||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY    803   TATCTGCGGTFACGTCGGCCCTGTGTTTATTGTTGATGCCCATCTGTTGCAG 862
      ||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB    282450 TACTTCCTGTCCTGCTTGATGGAGCTTACTCGCTATAGGCGATTTCATGCCGTAATG 282509
      ||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY    863   GGATGGTGAGATTTCGGGACCCGCAGTAGATTGC 900
      ||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB    282510 GGAGTCGGCAAGTCCTCGTCTCGCTTGAGAGTTTC 282547
      ||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

Search completed: July 6, 2003, 06:07:24  
Job time : 3761 secs



P: Brädon W, Naughton A, Robins K, Shaw N, Tinschert A;

PI Zimmermann I;

XX WPI: 1998-101063/29.

DR P-PSDB: AAW40263.

XX Tri-fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation

PT - by stereoselective hydrolysis of corresponding racemic anide using

PT microorganism or derived enzyme, used as drug intermediate

XX Claim 7a: Page 32-34; 68pp; German.

PS This DNA sequence encodes a R-specific amidohydrolase isolated from

CC Klebsiella oxytoca strain PRS1 which allows the microorganism to

CC utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole

CC nitrogen source. This amidohydrolase is used in a process for preparing

CC (R)-isomers of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which

CC is cheaper than prior art optical resolution of the racemate using

CC dimethoxy strychnine or (S)-(-)-alpha-methylbenzylamine.

XX Sequence 1442 BP; 365 A; 350 C; 370 G; 337 T; 0 other;

SQ

Query Match 100.0% Score 1442; DB 19; Length 1442;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCGGAACTCCATGTGGCGGATGCTCTGGTCGAGCAGGATATCGCATGATCCAGCGG 50

DB 1 CCCGGAACTCCATGTGGCGGATGCTCTGGTCGAGCAGGATATCGCATGATCCAGCGG 60

QY 61 CGGCACAGCGGTGGCGGTAAATGGATAAAGCCCTGGTGTAGAAACGCTGACCCCAAC 120

DB 61 CGGCACAGCGGTGGCGGTAAATGGATAAAGCCCTGGTGTAGAAACGCTGACCCCAAC 120

QY 121 AGCTCTGTGATGATCTTTAATAGCTGCTCATCTGCTCTGTAACGCTATATAAT 180

DB 121 AGCTCTGTGATGATCTTTAATAGCTGCTCATCTGCTCTGTAACGCTATATAAT 180

QY 181 AGTGGAGAATAACATATGAATGGTGGAGAATCCATTATGCCAAAGCGGTGG 240

DB 181 AGTGGAGAATAACATATGAATGGTGGAGAATCCATTATGCCAAAGCGGTGG 240

QY 241 TCCCGGGGTAAACCGGTAAAGCGATCCTCAGCGAAGAATGCAAAAGATTTCATTA 300

DB 241 TCCCGGGGTAAACCGGTAAAGCGATCCTCAGCGAAGAATGCAAAAGATTTCATTA 300

QY 301 CACCATTTGGCTTATTCACACCGCTGCTGACCATGGAACCGGTCAGCGGATATTGT 360

DB 301 CACCATTTGGCTTATTCACACCGCTGCTGACCATGGAACCGGTCAGCGGATATTGT 360

QY 361 CGACATCGAGATGCTTTTGAGGTGCTATCAATCGGACAGCATATTCGAGCGAGT 420

DB 361 CGACATCGAGATGCTTTTGAGGTGCTATCAATCGGACAGCATATTCGAGCGAGT 420

QY 421 GCTAAATATGCTTCTCAACCGCAAAACGACCGATATGCTCAATGCGCGGAGAA 480

DB 421 GCTAAATATGCTTCTCAACCGCAAAACGACCGATATGCTCAATGCGCGGAGAA 480

QY 481 AGGTGATGCTGCTGCTTATCGAATCCATGTTGCGCGCGGCTGATCCGTAAGG 540

DB 481 AGGTGATGCTGCTGCTTATCGAATCCATGTTGCGCGCGGCTGATCCGTAAGG 540

QY 541 CATCTGCGCCATGATTCGCGATTTGGCGGATGACCGGACCGCTGAGCGGATGCT 600

DB 541 CATCTGCGCCATGATTCGCGATTTGGCGGATGACCGGACCGCTGAGCGGATGCT 600

QY 601 CAATGATCGCTGCCAGAAAGGTGCGCATGATTAAACTCCACAGTGAAGAGTCTACTG 660

DB 601 CAATGATCGCTGCCAGAAAGGTGCGCATGATTAAACTCCACAGTGAAGAGTCTACTG 660

QY 661 GAGCAACGCGATTCGCTTATTAACCCCATATTGGCACCTTGGACGCTATGGCCAGA 720

DB 661 GAGCAACGCGATTCGCTTATTAACCCCATATTGGCACCTTGGACGCTATGGCCAGA 720

QY 721 AATTGACTCAATCAATCTGACCTCAGACAATACGGCGGGGAATATGGAATGTGCGCGA 780

DB 721 AATTGACTCAATCAATCTGACCTCAGACAATACGGCGGGGAATATGGAATGTGCGCGA 780

QY 781 TATAGNCCAGGAGATATACCTATCTCCGCTAGCTGCGCTCGAGGCGCGCTGTTAT 840

DB 781 TATAGNCCAGGAGATATACCTATCTCCGCTAGCTGCGCTCGAGGCGCGCTGTTAT 840

QY 841 TGGTGTATGCCATGCTTGTGAGGATGATGGTATTTGCGGGACCGCAGTAGATTTGC 900

DB 841 TGGTGTATGCCATGCTTGTGAGGATGATGGTATTTGCGGGACCGCAGTAGATTTGC 900

QY 901 CTCATATCACCACCATCAAGTCGATTTGATCAAGAACTGGCAGCTTTCTTGGCCACGAAT 960

DB 901 CTCATATCACCACCATCAAGTCGATTTGATCAAGAACTGGCAGCTTTCTTGGCCACGAAT 960

QY 961 GGAGAAATGCCGAAATATATGAGTATTTGGCAGTGCAGCTCCGCTGGAGGATGCGACGG 1020

DB 961 GGAGAAATGCCGAAATATATGAGTATTTGGCAGTGCAGCTCCGCTGGAGGATGCGACGG 1020

QY 1021 AATTGCAATATCGGACTTAATTTACTGCTGATGAGAGACTTTGGCTTCGAACAATGGGA 1080

DB 1021 AATTGCAATATCGGACTTAATTTACTGCTGATGAGAGACTTTGGCTTCGAACAATGGGA 1080

QY 1081 TGCCTACATGCTTCTGAGTCAATGCGCAAAATGCGGCTGGCAACATGCTGACCCCAA 1140

DB 1081 TGCCTACATGCTTCTGAGTCAATGCGCAAAATGCGGCTGGCAACATGCTGACCCCAA 1140

QY 1141 ATACACCGTTGGCGGATGCTGAAACAAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200

DB 1141 ATACACCGTTGGCGGATGCTGAAACAAAACCTGTTAGTTAGTAGGAATAACTAACCGG 1200

QY 1201 TGAACATTAACCGGATGATGATCGGGTAAATGCTGAGTTCACAACTCGCTATTTAA 1260

DB 1201 TGAACATTAACCGGATGATGATCGGGTAAATGCTGAGTTCACAACTCGCTATTTAA 1260

QY 1261 CAGCTAAAGCAGTGCATATGCGGCGCAGATACACCAATCAATATGTTTACTTCTCC 1320

DB 1261 CAGCTAAAGCAGTGCATATGCGGCGCAGATACACCAATCAATATGTTTACTTCTCC 1320

QY 1321 TTAACGAGTGCAGGCGCAGATGTCACAAATGCGGCGGAGCAACCCAGCTATT 1380

DB 1321 TTAACGAGTGCAGGCGCAGATGTCACAAATGCGGCGGAGCAACCCAGCTATT 1380

QY 1381 GCGGAAATTAATCAAAATGCGGCGATCAAGGCGACCACTCAATGCAATTCATTGGAT 1440

DB 1381 GCGGAAATTAATCAAAATGCGGCGATCAAGGCGACCACTCAATGCAATTCATTGGAT 1440

QY 1441 CC 1442

DB 1441 CC 1442

RESULT 2

AA264148

ID AA264148 standard; DNA: 945 BP.

XX AA264148;

AC AA264148;

CT 23 FEB-2001 (first entry)

XX

TF Enterobacter cloacae stereoselective amidase DNA coding sequence.

XX Amidase: stereoselective hydrolysis; alpha-amino acid amide;

XX alpha-hydroxy acid amide; optically active product;

XX recombinant production; ds.

XX Enterobacter cloacae.

XX W0200063354-A1.

XX 26-OCT-2000.





```

FT misc_feature 300001..349980
FT /note= b
FT /tag= a
FT AAH41225"
XX
PN FR292651-A1.
XX
XX 27-OCT-2000.
XX
XX 21-APR-1999; 99PR-0005034.
XX
XX 21-APR-1999; 99PR-0005034.
XX
PA (CNRS ) CNRS CENT NAT RICH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI: 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry .
XX
XX Claim 1: Page 347-443; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
XX hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX vents. The present sequence is a fragment of the genomic sequence of P.
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAB41223
XX and the 3' end of this sequence overlaps with the 5' end of AAB41225. The
XX proteins of the present invention have various potential industrial uses,
XX since the proteins are stable at very high temperatures, some up to 110
XX degrees centigrade.
XX Note: This patent is in the same patent family as WO2000065052, which
XX contains additional sequences as shown in AAB99132-AAB99143.
XX AAH75903-AAH75920 and AAH56436.
XX
XX Sequence 349980 BP; 94090 A; 78692 C; 78319 G; 98879 T; 0 other;
XX
XX Query Match 2.9%; Score 41.2; DB 22; Length 349980;
XX Best Local Similarity 53.8%; Pred. No. 0.39;
XX Matches 85; Conservative 0; Mismatches 73; Indels 6; Gaps 6;
XX
XX 743 ACCCCAGCAATCACGGCGGGAATATGGATGTCCCGATATAGGACCGAGGATATTACC 802
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 282390 ACCCCAGGAACACGGCGGGAACATGGACACAAACCTAATCAGAAAGGGAACACCAATT 282449
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 803 TATCTCCGGTACGTCGGCTCGAGCCGCCCTGTTTATTGTGATGCCCATGCTGTGTCAG 862
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 282450 TACTTCCCTGCTCTCTTGTATGAGAGCTTACCTCGGATAGGCGATTTCATGCCGTAATG 282509
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 863 GGTGATGGTGAGATTTCGGCGACCGCAGTAGAGATTTC 900
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 282510 GGAGATGGCGAAGCTGCGTCTCGGCTGTGAGGTTTC 282547
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
RESULT 5
AAQ66950/c
ID AAQ66950 standard; cDNA; 3900 BP.
XX
XX AAQ66950;
XX
XX 15-FEB-1995 (first entry);
XX
XX Cell death hid cDNA.
XX
XX Cell death; head involution defective; hid gene; reaper gene; rpr;
XX apoptosis; transgenic animal; antisense; ss.
XX
XX Drosophila melanogaster.
XX

```

```

XX FH Key location/Qualifiers
XX FT 450..1632
XX FT /*tag= a
XX
XX DN W09416071-A.
XX
XX XX 21-JUL-1994.
XX
XX XX 14-JAN-1994; 94WO-US00500.
XX
XX XX 15-JAN-1993; 93US-0004957.
XX
XX XX 17-SEP-1993; 93US-0123343.
XX
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Abrams JM, Grether ME, Steller H, White K;
XX WPI: 1994-249218/30.
XX P-PSDB; AAR55791.
XX
XX Isolated cell death genes from Drosophila - and novel assays for
XX apoptotic cell deaths and apoptotic and necrotic cell deaths,
XX based on selective staining by toluidine blue, acridine orange
XX and Nile blue
XX
XX Disclosure: Fig.6; 93pp; English.
XX
XX The reaper (rpr) and head involution defective (hid) genes, mapping
XX to position 75c1.2 of D. melanogaster chromosome 3, exhibit
XX expression patterns related to the pattern of cell death during
XX Drosophila embryogenesis. Cell death genes, or antisense sequences,
XX can be used to reduce or abolish apoptosis, e.g. in transgenic
XX animals. An hid cDNA sequence is given in AAQ66950, and the
XX deduced sequence of the encoded HID protein in AAR55791.
XX
XX Sequence 3900 BP; 1288 A; 876 C; 736 G; 1000 T; 0 other;
XX
XX Query Match 2.7%; Score 39.2; DB 15; Length 3900;
XX Best Local Similarity 49.1%; Pred. No. 0.11;
XX Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
XX
XX 320 ACACCCGTCCTGACCATCGAACCCGCTACCCGGATTATGTCGACACGAGATGCTTTT 379
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 1458 ACACCCGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1399
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 380 GAAGGTGCTATCAATTCGCAACAGGATATTCGAGCCAGTTCCTAAATATGCCCTTCTC 439
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 1398 GGAGCTCTCTTCTTCTCCGGGGGGGAGTGGTCCATTGACTCTCGACACGCTCCGGCC 1339
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 440 AACCCACAAAACGAGCGATCATGTCTAATGCGCGGAGAAAGGTGATGTCGCTGCTC 499
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 1338 GGCGGTAAGTTCGCTAGCGATCGCAAACTCGTCCCAAGTGGCGTCAATGATCGCTG 1279
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 500 TATATGAATCGATGTTGCGCGCGCGCGGTGA 531
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX 1276 GACTCCGCTCATCTCCCGCGCGGAGGTCA 1247
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
RESULT 6
AAR08913/c
ID AAR08913 standard; cDNA; 3988 BP.
XX
XX ABL08913;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 21221.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX

```

```
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB64810.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 21221; 21pp - Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX Sequence 3988 BP; 1316 A; 893 C; 753 G; 1026 T; 0 other;

Query Match 2.7%; Score 39.2; DB 23; Length 19867;
Best Local Similarity 49.1%; Pred. No. 0.3;
Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 320 ACACCCGCTCTGACCATCGAACCCGGTGACCGGATTATTGTCGACATCGAGATGCTTT 379
DB 1545 ACCACCGGGCTGGCTATCGGTATGACACACGGATTATTGCTGCTCGAGTGGCTATT 1486
QY 380 GAAGTGCTATCAATTCGGAACAGGATATTCGAGCCAGTTCGTAATAATGCCCTTTTC 439
DB 1485 GGAGCTCTTCTTTCGGGGGGGGATCGCTCCATTGAACCTCTCGACAGCGCTCCCGGCC 1426
QY 440 AACCCACAAAACGGACCATGATGTTCAATGGCGGGAGAAAGTGATGTCGCTGCTGTC 499
DB 1425 GGCGGTAAGTTGCTAGCGATCGCAAACTCGTCCCAAGTGGCTCATGATGCTGTCIG 1366
QY 500 TATATCGAATCATGTTGCCCGCGCGGCTTGA 531
DB 1365 GTACTCGGCTCATCTCGCGCGTTCGAGGTCA 1334

RESULT 7
ABL06912/c
ID ABL06912 standard; cDNA; 19867 BP.
XX ABL06912;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 21218.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
XX
```

```
PN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB64809.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 21218; 21pp - Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX Sequence 19867 BP; 5006 A; 3967 C; 3859 G; 6035 T; 0 other;

Query Match 2.7%; Score 39.2; DB 23; Length 19867;
Best Local Similarity 49.1%; Pred. No. 0.3;
Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 320 ACACCCGCTCTGACCATCGAACCCGGTGACCGGATTATTGTCGACATCGAGATGCTTT 379
DB 14456 ACCACCGGGCTGGCTATCGGTATGCGACACTGGATTATTGCTGCTCGAGTGGCTATT 14397
QY 380 GAAGTGCTATCAATTCGGAACAGGATATTCGAGCCAGTTCGTAATAATGCCCTTTTC 439
DB 14396 GGAGCTCTTCTTTCGGGGGGGGATCGCTCCATTGAACCTCTCGACAGCGCTCCCGGCC 14337
QY 440 AACCCACAAAACGGACCATGATGTTCAATGGCGGGAGAAAGTGATGTCGCTGCTGTC 499
DB 14336 GGCGGTAAGTTGCTAGCGATCGCAAACTCGTCCCAAGTGGCTCATGATGCTGTCIG 14277
QY 500 TATATCGAATCAATGTTGCCCGCGCGGCTTGA 531
DB 14276 GTACTCGGCTCATCTCGCGCGTTCGAGGTCA 14245

RESULT 8
AAA8:477/c
ID AAA8:477 standard; DNA; 26778 BP.
XX AAA8:477;
XX 04-DEC-2000 (first entry)
XX N. meningitidis partial DNA sequence gnm_25 SEQ ID NO:25.
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX Meningococcus B; MenB; ds.
XX Neisseria meningitidis.
XX
```

```
FN WO20022430-A2.
XX
XX
PD 20-APR-2000.
XX
XX 06-OCT-1999; 99WO-US23573.
XX
XX 09-OCT-1998; 98US-0103794.
XX
XX 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,
PI Rappuoli R, Pizza M;
XX
XX WPI: 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PI used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea.
XX
XX Claim 7; Page 524-531; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
XX Sequence 26778 BP; 6602 A; 7308 C; 5846 G; 6313 T; 4 other;
SQ
Query Match 2.7%; Score 39.2; DB 21; Length 26778;
Best Local Similarity 52.4%; Pred. No. 0.37;
Matches 86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 188 GAATAACATATGAATGCTTGAGAGATCCATTATGCGCAACCGCGTGTGTGCGCGG 247
Db 23055 GCATCTTAAACAAAACACGCAAAAGCTATGATGACCGGAAGAGAGTATACCGCGG 22994
QY 248 CGTAAACGGTAAACGATCATCTGACGAGAGAAATGCAAAAGAGTTTCATACACATT 307
Db 22995 AGCGAGGACATTATTGTACACATCGAAGACCAATATCGAGCAAGCGCTGGGATTGAC 22936
QY 308 GGCCCTTATTCACACCGCTGCGTACCATCATGACCAACCGCGTGACCG 351
Db 22935 TGCCGGTTTCCGGAAGAGTCTTGACATCGAAGCGGAGACGG 22892
RESULT 9
AAF21612
ID AAF21612 standard; DNA: 349960 BP.
XX
AC AAF21612;
```

```
XX
XX 13-MAR-2001 (first entry)
XX
XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
XX diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX ds.
XX
XX Neisseria meningitidis.
XX
XX WO2000064791-A1.
XX
XX 09-NOV-2000.
XX
XX 05-MAR-2000; 2000WO-US05928.
XX
XX 30-APR-1999; 99US-0132068.
XX
XX 08-OCT-1999; 99WO-US23573.
XX
XX 28-FEB-2000; 2000GB-0004695.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V,
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R,
PI Frazer CM, Grandi G;
XX
XX WPI: 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections.
XX
XX Claim 7; Appendix A: 692pp; English.
XX
XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21589 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
XX Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;
SQ
Query Match 2.7%; Score 39.2; DB 21; Length 349980;
Best Local Similarity 52.4%; Pred. No. 1.9;
Matches 86; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 189 GAATAACATATGAATGCTTGAGAGATCCATTATGCGCAACCGCGTGTGTGCGCGG 247
Db 275981 GCATCTTAAACAAAACACGCAAAAGCTATGATGACCGGAAGAGTATTAICCGCGG 276040
QY 248 CGTAAACCGGTAACGATCATCTGACGAGAGAAATGCAAAAGAGTTTCATACACATT 307
Db 276041 AGCGAGGACATTATTGTACACATCGAAGACCAATATCGAGCAAGCGCTGGGATTGAC 276100
QY 308 GGCCCTTATTCACACCGCTGCGTACCATCATGACCAACCGCGTGACCG 351
```





```
CC thought to inhibit Hid induced cell death. Mutant forms of the hid
CC gene, hidA1a3 and hidA1a5 have been generated in which DNA encoding
CC phosphoacceptor residues is replaced with DNA encoding
CC non-phosphorylatable amino acids. These mutants are not suppressed
CC by MAPK induced phosphorylation. Compositions from the mutant gene
CC are useful in screening assays for antagonists and agonists of Hid
CC phosphorylation by mitogen activated protein kinase (MAPK). They may
CC also be used to identify new constituents of the Hid signalling
CC pathway. The new compositions may be useful in identifying new homologs
CC of Hid or its natural mutations. The present sequence is the cDNA
CC of the hidA1a3 mutant.
XX
XX
SQ Sequence 1233 BP; 247 A; 453 C; 330 G; 203 T; 0 other;

Query Match 2.6%; Score 37.6; DB 21; Length 1233;
Best Local Similarity 48.6%; Pred. No. 0.18;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 320 ACACCGCTCTCAGCATGAAACCGGTGACCGGATTATTGTGACACATCGAGATGCTTT 379
DB 1059 ACCACCGGACTGCGTGGTATGCGACACTGAGTATTGCTGCTGAGTGGCTATT 1000
QY 380 GAAGTGTCTATCAATTCGGAACAGATATTCGACCCAGTTGCTAATAATGCGCTTC 439
DB 999 GGAGCTCTCTCTTCCTCCGGCGGGAIGGTCATTGAACCTCTGACAGACTCCCGGCC 940
QY 440 AACCCACAAACGGACCGATCATGTCATGCGCGGGAAGGATGATCTGCTGCTGTC 499
DB 939 GSCCGTAAAGTGTGCTAGCAATGCCAACTCGTCCCAAGTGGCTGATGATGCTG 880
QY 500 TATATCGAATCATTTGCTCCCGCGCGGTGA 531
DB 879 GTACTCGGCTCATCTCCGCCGTCGAGGTCA 848

RESULT 14
AAA27104/C
ID AAA27104 standard; cDNA; 1233 BP.
XX
XX AAA27104;
XX
XX 31-JUL-2000 (first entry)
XX
XX Drosophila melanogaster hidA1a5 cDNA.
XX
XX Fruit fly; Hid; mitogen activated protein kinase p42/p44;
XX MAPK; phosphorylation; apoptosis; ss.
XX
XX Drosophila melanogaster.
XX Synthetic
XX
XX Key Location/Qualifiers
XX CDS 1..1233
XX FT FT /*tag= a
XX FT FT /product= HidA1a5 protein
XX FT FT 361..362
XX FT FT /*tag= b
XX FT FT /label= wild type A G replaced with G C
XX FT FT 442
XX FT FT /*tag= c
XX FT FT /label= Wild type A replaced with G
XX FT FT 536
XX FT FT /*tag= d
XX FT FT /label= Wild type A replaced with G
XX FT FT /*tag= e
XX FT FT /label= Wild type A replaced with G
XX FT FT 751..752
XX FT FT /*tag= f
XX FT FT /label= wild type A G replaced with G C
XX
XX WO200026634-A2.
XX
XX
```

```
PD 11-MAY-2000.
XX
XX 29-OCT-1999; 99WO-US25514.
XX
XX 29-OCT-1998; 98US-0106108.
XX
XX 29-OCT-1999; 99US-0431573.
XX
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Steller H, Agapite J, Bergmann A, McCall K;
XX
XX WP; 2000-376155/32.
XX
XX P-203; RAY94424.
XX
XX Composition, useful in screening assays for antagonists and agonists of
XX Hid phosphorylation by mitogen activated protein kinase (MAPK).
XX comprises DNA having an oligonucleotide sequence of mutant forms of the
XX Drosophila hid gene -
XX
XX Claim 1; Fig 11; 58pp; English.
XX
XX The product of the Drosophila melanogaster hid gene is responsible for
XX inducing apoptosis in embryos. Activation of the Ras/MAPK pathway is
XX thought to inhibit Hid induced cell death. Mutant forms of the hid
XX gene, hidA1a3 and hidA1a5 have been generated in which DNA encoding
XX phosphoacceptor residues is replaced with DNA encoding
XX non-phosphorylatable amino acids. These mutants are not suppressed
XX by MAPK induced phosphorylation. Compositions from the mutant gene
XX are useful in screening assays for antagonists and agonists of Hid
XX phosphorylation by mitogen activated protein kinase (MAPK). They may
XX also be used to identify new constituents of the Hid signalling
XX pathway. The new compositions may be useful in identifying new homologs
XX of Hid or its natural mutations. The present sequence is the cDNA
XX of the hidA1a5 mutant.
XX
XX Sequence 1233 BP; 245 A; 454 C; 331 G; 203 T; 0 other;

Query Match 2.6%; Score 37.6; DB 21; Length 1233;
Best Local Similarity 48.6%; Pred. No. 0.18;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 320 ACACCGCTCTCAGCATGAAACCGGTGACCGGATTATTGTGACACATCGAGATGCTTT 379
DB 1059 ACCACCGGACTGCGTGGTATGCGACACTGAGTATTGCTGCTGAGTGGCTATT 1000
QY 380 GAAGTGTCTATCAATTCGGAACAGATATTCGACCCAGTTGCTAATAATGCGCTTC 439
DB 999 GGAGCTCTCTCTTCCTCCGGCGGGAIGGTCATTGAACCTCTGACAGACTCCCGGCC 940
QY 440 AACCCACAAACGGACCGATCATGTCATGCGCGGGAAGGATGATGCTGCTGCTGTC 499
DB 939 GSCCGTAAAGTGTGCTAGCAATGCCAACTCGTCCCAAGTGGCTGATGATGCTG 880
QY 500 TATATCGAATCATTTGCTCCCGCGCGGTGA 531
DB 879 GTACTCGGCTCATCTCCGCCGTCGAGGTCA 848

RESULT 15
ABL20343
ID ABL20343 standard; DNA; 4810 BP.
XX
XX ABL20343;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 12502.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw mode:

Run on: July 6, 2003, 04:47:45 ; Search time 97 Seconds  
(without alignments)  
4559.050 Million: cell updates/soc

Title: US-10-086-082-1

Perfect score: 1442  
Sequence: 1 cccgggaactccatgagcc.....aatgaattcattggatcc 1442

Scoring table: IDENTITY\_NJC  
Gapop 10.0 , Gapext 1.0

Searched: 441462 seqs, 153338361 residues

Total number of hits satisfying chosen parameters: 38272;

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA:\*

1: /cgn2\_6/plodata/1/ina/5A\_COMB.seq:\*

2: /cgn2\_6/plodata/1/ina/5A\_COMB.seq:\*

3: /cgn2\_6/plodata/1/ina/5A\_COMB.seq:\*

4: /cgn2\_6/plodata/1/ina/65\_COMB.seq:\*

5: /cgn2\_6/plodata/1/ina/PCFUS\_COMB.seq:\*

6: /cgn2\_6/plodata/1/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37.5	2.5	501	1 US-08-123-343A-15	Sequence 15, Appli
C 2	37.6	2.5	1233	4 US-09-431-573-1	Sequence 1, Appli
C 3	37.6	2.6	1233	4 US-09-431-573-2	Sequence 2, Appli
C 4	37.6	2.6	1233	4 US-09-431-573-3	Sequence 3, Appli
C 5	37.6	2.6	3500	1 US-08-123-343A-6	Sequence 6, Appli
C 6	35.2	2.4	428	4 US-09-144-367-7	Sequence 7, Appli
C 7	34.2	2.4	5733	2 US-08-473-558A-1	Sequence 1, Appli
C 8	33.2	2.3	3201	2 US-08-633-770A-3	Sequence 3, Appli
C 9	33.2	2.3	4726	1 US-06-633-770A-11	Sequence 11, Appli
C 10	32.8	2.3	7218	1 US-08-232-463-14	Sequence 14, Appli
C 11	32.4	2.2	3376	4 US-09-309-572-3	Sequence 3, Appli
C 12	32.4	2.2	9551	1 US-08-056-200-93	Sequence 93, Appli
C 13	32.4	2.2	9551	2 US-08-800-644-93	Sequence 93, Appli
C 14	32.2	2.2	1226	1 US-08-374-688-4	Sequence 4, Appli
C 15	32.2	2.2	1349	1 US-08-374-688-1	Sequence 1, Appli
C 16	32.2	2.2	6254	4 US-08-927-219-126	Sequence 126, App
C 17	31.8	2.2	1817	4 US-08-943-731-193	Sequence 193, App
C 18	31.8	2.2	20084	4 US-08-943-731-5	Sequence 5, Appli
C 19	31.6	2.2	1380	4 US-09-513-783A-169	Sequence 169, App
C 20	31.2	2.2	2157	2 US-08-743-637B-16	Sequence 16, Appli
C 21	31.2	2.2	2157	3 US-08-526-840B-15	Sequence 15, Appli
C 22	31	2.1	4403765	4 US-09-103-840A-2	Sequence 2, Appli
C 23	31	2.1	4411529	4 US-09-103-840A-1	Sequence 1, Appli
C 24	30.8	2.1	1187	5 PCT-US95-13536-2	Sequence 2, Appli
C 25	30.8	2.1	1188	5 PCT-US95-13536-1	Sequence 1, Appli
C 26	30.6	2.1	1542	4 US-09-202-894B-3	Sequence 3, Appli
C 27	30.6	2.1	4403765	4 US-09-103-840A-2	Sequence 2, Appli

C 28	30.6	2.1	4411529	4 US-09-103-840A-1	Sequence 1, Appli
C 29	30.4	2.1	1049	3 US-08-682-643-1	Sequence 1, Appli
C 30	30.2	2.1	747	4 US-08-998-416-1016	Sequence 1016, Ap
C 31	30.2	2.1	10095	3 US-08-822-586-45	Sequence 45, Appli
C 32	30	2.1	1614	4 US-09-522-217-7	Sequence 7, Appli
C 33	30	2.1	2665	3 US-09-040-005-1	Sequence 1, Appli
C 34	29.8	2.1	289	4 US-09-007-005-17	Sequence 17, Appli
C 35	29.8	2.1	289	4 US-09-244-796-17	Sequence 17, Appli
C 36	29.8	2.1	417	4 US-08-936-165A-17	Sequence 17, Appli
C 37	29.4	2.0	1888	1 US-08-485-107-1	Sequence 1, Appli
C 38	29.4	2.0	7802	4 US-08-961-527-112	Sequence 112, App
C 39	29.2	2.0	1524	4 US-09-134-001C-2055	Sequence 2055, Ap
C 40	29.2	2.0	46899	1 US-08-471-119A-1	Sequence 1, Appli
C 41	29	2.0	300	1 US-08-083-946-6	Sequence 6, Appli
C 42	29	2.0	300	3 US-08-452-915-6	Sequence 6, Appli
C 43	29	2.0	615	4 US-09-134-001C-138	Sequence 138, App
C 44	29	2.0	850	1 US-07-611-528A-1	Sequence 1, Appli
C 45	29	2.0	850	1 US-08-083-946-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-123-343A-15/C  
Sequence 15, Application US/08123343A  
Patent No. 5593879  
GENERAL INFORMATION:  
APPLICANT: Steller, Hermann  
APPLICANT: Abrams, John M.  
APPLICANT: Grether, Megan F.  
TITLE OF INVENTION: Cell Death Genes of Drosophila  
TITLE OF INVENTION: Melanogaster and Vertebrate Analogs  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/123,343A  
FILING DATE: 17-SEP-1993  
CLASSIFICATION: 80C  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/004,957  
FILING DATE: 15-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5907A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 61861-9540  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-123-343A-15

Query Match 2.6%; Score 37.6; DB 1: Length 503;  
Best local Similarity 48.6%; Pred. No. 0.02;  
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;





```
Db 939 GCGGTAAAGTGTCTGATAGCATCGCAAAACGCTCTCCAACTGCGCTTCATGATGCTCTG 880
Qy 500 TATATCGAATCCATGTTGCCCGCGGGCTTCA 532
Db 879 GTACTCGGCTCATCCCGCGCGTGGAGTCA 848

RESULT 5
US-08-123-343A-6/c
: Sequence 6, Application US/08123343A
: Patent No. 5593879
: GENERAL INFORMATION:
: APPLICANT: Steller, Herman;
: APPLICANT: Abrams, John M.;
: APPLICANT: Grether, Megan E.;
: APPLICANT: White, Kristin
: TITLE OF INVENTION: Cell Death Genes of Drosophila
: TITLE OF INVENTION: Melanogaster and Vertebrate Analogs
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: US
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/123.343A
: FILING DATE: 17-SEP-1993
: CLASSIFICATION: 50C
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/004,957
: FILING DATE: 15-JAN-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: MIT-5907A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 61861-9540
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3900 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
US-08-123-343A-6

Query Match 2.6%; Score 37.6; DB 1; Length 3900;
Best Local Similarity 48.68; Pred. No. 0.072;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 320 ACACCCGCTCGACCATCGAACCCGGTACCGGATTTATTCGACACTCGAGATGTTT 379
Db 1458 ACCACCGGACTGGCTGCGGTATGGCAGACTGGATTTATTCGCTGCTCGAGTGGCTATT 1399
Qy 380 GAAGGTGCTATCAATTCGGACACAGATATCCGAGCCAGTTGCTAAAATGCGCTTTC 439
Db 1398 GGAGCTCTTCTTCGCGGGCGGATGCGTCCATTTGAACCTCTTCGACAGCGCTCCCGCC 1339
Qy 440 AACCCACAAACGGACCGCATCGTCAATGGCCGGAGAAAGTGAATGTCGCTGCTC 499
Db 1338 GCGGTAAAGTGTCTGATAGCATCGCAAACTCGTCCCAAGTGGCTTCATGATGCTG 1279
Qy 500 TATATCGAATCCATGTTGCCCGCGGGCTTCA 531
Db 1278 GTACTCGGCTCATCTCGCGCGTGGAGTCA 1247
```

```
RESULT 6
US-09-144-367-7/c
: Sequence 7, Application US/09144367
: Patent No. 6432639
: GENERAL INFORMATION:
: APPLICANT: Lichter, Jay
: APPLICANT: Guido, Marco
: TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
: FILE REFERENCE: SEQ-12P
: CURRENT APPLICATION NUMBER: US/09/144,367
: CURRENT FILING DATE: 1998-08-31
: PRIOR APPLICATION NUMBER: 60/058,612
: PRIOR FILING DATE: 1997-09-10
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 7
: LENGTH: 429
: TYPE: DNA
: ORGANISM: H. sapiens
: FEATURE:
: NAME/KEY: Other
: LOCATION: (0)...(0)
US-09-144-367-7

Query Match 2.4%; Score 35.2; DB 4; Length 429;
Best Local Similarity 57.5%; Pred. No. 0.12;
Matches 61; Conservative 1; Mismatches 44; Indels 0; Gaps 0;

Qy 1164 ACAAAACCGTGTAGTTTAGTAGGATACTAACCGGTGAACATTACCCGGATGTAGATC 1223
Db 315 ACACATATCTTCAATGTACTCAAACTACATGACTGATATTTTAAGTGGATGAATAC 256
Qy 1224 GCGTAATGTGTAAGTTCACAAATCGCTATTTTAACACAGCTAAAG 1269
Db 255 ATGGGATTTATATCTCATATAAGCAGTTATTTTAAGAGAGMAAG 210

RESULT 7
US-09-473-553A-1
: Sequence 1, Application US/08473553A
: Patent No. 5859338
: GENERAL INFORMATION:
: APPLICANT: Meyerowitz, Elliot M.
: APPLICANT: Clark, Steven E.
: APPLICANT: Williams, Robert W.
: TITLE OF INVENTION: Plant Clavatal Nucleic Acids,
: TITLE OF INVENTION: Transformed Plants, and Proteins
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fleer, Holbach, Test, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/473,553A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Silva, Robin M.
: REGISTRATION NUMBER: 38,304
: REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
```

```
TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5733 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2434..5037
: NAME/KEY: CDS
: LOCATION: 5117..5467
: US-08-473-553A-1

Query Match      2.4%; Score 34.2; DB 2: Length 5733;
Best Local Similarity 50.3%; Pred. No. 1.2;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Oy 1136 CCCAATATACACCGTTGGCGGATGCTGAACAAACCTGTAGTTAGTAAAGTA 1195
Db 3709 CTCAACGGGACATGTTCCGGGGGGCTTTTCATCTACCGGTAGTTACGATTG 3768
Oy 1196 ACCGGTGAACATACCCGGATGTAGTCGGGTAAAGTAAAGTTCAACAAAGCTATT 1255
Db 3769 ACTGATAATTCTCTCCGGTGAACCTCCGGTAACGATGTCGGGGATGTTCTCGATCAG 3828
Oy 1256 TTAAACAGCTAAAGCAGGTGCATATGCGGCGCAGATACACCCATCAAT 1302
Db 3829 ATTATCTCTTAACAACTGTTTCCGGGGAGATTCACCTGGCAT 3875

RESULT 8
US-08-633-770A-3/c
: Sequence 3, Application US/08633770A
: Patent No. 5903760
: GENERAL INFORMATION:
: APPLICANT: Hojsten, Kirsten
: APPLICANT: Yu, Shukun
: APPLICANT: Kragh, Karsten
: APPLICANT: Christensen, Tove
: APPLICANT: Marcussen, Jan
: TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
: TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/633,770A
: FILING DATE: July 8, 1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/EP94/03398
: FILING DATE: OCT-15-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Altman, Daniel E
: REGISTRATION NUMBER: 34,115
: REFERENCE/DOCKET NUMBER: DY006.001APC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 714-760-0404
: TELEFAX: 714-760-9502
: TELEX:

: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4726 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA

: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3201 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-633-770A-3

Query Match      2.3%; Score 33.2; DB 2: Length 3201;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Oy 243 CGCGGGTAAACGGGTAAAGCATCACTGACCGGAAGAAATGCAAAAAGAGTTTCATTA 302
Db 1628 CTGGGCTTCACTTTCAGACGTCACACAGACCTGTATAAGTCGTGACATTTCCGGGACC 969
Oy 303 CCATGGGCCCTATTCCACACCCGGTCTGACCATCGAACCCGGTGACCGGATTAATGTCG 362
Db 968 GTATCGGCACTGATACCGTAACAAATCGATACCAACCATACCTGCTCCAGCTTGATGAT 909
Oy 363 ACATCCAGATGCTTTTGAAGGTGCTATCAATTCGAAACAGGATAT 408
Db 908 CCGAGTTGGCTTCCAAAGTCGATGGCAATTTGAGAGTASTTAT 863

RESULT 9
US-08-633-770A-11/c
: Sequence 11, Application US/08633770A
: Patent No. 5908760
: GENERAL INFORMATION:
: APPLICANT: Hojsten, Kirsten
: APPLICANT: Yu, Shukun
: APPLICANT: Kragh, Karsten
: APPLICANT: Christensen, Tove
: APPLICANT: Marcussen, Jan
: TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
: TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/633,770A
: FILING DATE: July 8, 1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/EP94/03398
: FILING DATE: OCT-15-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Altman, Daniel E
: REGISTRATION NUMBER: 34,115
: REFERENCE/DOCKET NUMBER: DY006.001APC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 714-760-0404
: TELEFAX: 714-760-9502
: TELEX:

: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4726 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
```

US-08-633-770A-1:

Query Match 2.3%; Score 33.2; DB 2; Length 4726;  
Best Local Similarity 50.6%; Pred. No. 2.4;  
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

OY 243 CCGCGCTTAACCGGTACCGCATACCCGACGGGAGAAATGCAAAAAGAGTTTCAITACA 302  
DB 1741 CTGGGCTTCAACTTTTGAACGCTCCACAAAGAGCTGTATAAGTGTACAAATTCGGGGACC 1682  
OY 303 CCATGGCCCTTATCCACACCGCTGACCAATCGAACCCGGTACACGGGATTTATGTCG 362  
DB 1681 GTATCCGCACTAATCGGTAACTGATACCACTACCTGGTTCACAGGTTGATGATAG 1622  
OY 363 ACATCGAGATGCTTTTGAAGGTGCTATCAATTCGGAACAGATAT 408  
DB 1621 CCGTGGTGGCTTTTCCAAAGTCGATGCAATTTGAGAGTAGTTAT 1576

## RESULT 10

US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670467  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOMULPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0293

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent's Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/05/232.463  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/535.313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 1MMG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgt-Fls

US-08-232-463-14

Query Match 2.3%; Score 32.8; DB 1; Length 7218;  
Best Local Similarity 3.6%; Pred. No. 4.2;  
Matches 13; Conservative 193; Mismatches 160; Indels 0; Gaps 0;

OY 1051 GGTAGAACATTGGCTTCGAAACAATGGGATGCCTACATGCTTCTGAGTCAATGGGGCAA 1110

DB 1452 CATAGAAGAAATTCGGTACRR 1393  
OY 1111 AGTCGGCTCGGCAACATGTCGACCCCAATACACCGTTGGCGCGATGCTGAACAAAAA 1170  
DB 1392 RRR 1333  
OY 1171 CCGTTTGTAGTAGAATACTAACCGGTGAACATTAACCGGATGATAGATCGGGTAA 1230  
DB 1332 RRR 1273  
OY 1231 TGTGTAAGTTCAACAATCCCTATTTTAAACGTAAGCAGGTGCATATGGGCCAGAT 1290  
DB 1272 RRR 1213  
OY 1291 ACACCCATCAATATGTTTACTTACTCTTCAGCGGAGTGACGGCGGCAAGATGG 1350  
DB 1212 RRR 1153  
OY 1351 TCACATGGCGGAGCAACCCAGGCTATTCGCGAAATTAATCAAAATGGCGGCAACAAC 1410  
DB 1152 RRR 1093  
OY 1411 GGCAGA 1416  
DB 1092 RRRRRR 1097

## RESULT 11

US-05-309-572-3/c  
Sequence 3, Application US/05309572  
Patent No. 6440730  
GENERAL INFORMATION:  
APPLICANT: Heinrich-Pette-Institut  
TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV

FILE REFERENCE: P50489  
CURRENT APPLICATION NUMBER: US/05/309.572  
CURRENT FILING DATE: 1999-05-11  
EARLIER APPLICATION NUMBER: DE 198 56 463  
EARLIER FILING DATE: 1998-11-26  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 3476  
TYPE: DNA  
ORGANISM: Lymphocytic choriomeningitis virus  
US-05-309-572-3

Query Match 2.2%; Score 32.4; DB 4; Length 3376;  
Best Local Similarity 54.1%; Pred. No. 3.5;  
Matches 56; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 426 AAATGCGCTTTCACACCCCAACCAACCGGATCATGGTCAATGGCGGCGGAGAAAGGTG 485  
DB 2.41 AATAGCAATTTTCCACACAGACAGCGGAGTTCATACACTTTTACAGGAAACCCCTTG 2082  
OY 486 ATGTGCTCGCTGCTATAACGAATCCATGTGCCCCGGCGGTGATGCCATAGGCACT 545  
DB 2081 ATCAAAAACAATCAAGCAAGATTCGAATACACACGCGATGGATCTTTCGCGACCTCT 2022  
OY 546 GC 547  
DB 2021 TC 2020

## RESULT 12

US-08-056-200-93  
Sequence 93, Application US/08056200  
Patent No. 5616500  
GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu

APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-056-200-93

Query Match 2.2% Score 32.4; DB 1; Length 9551;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 3 CGGAACCTCCATGTGGCGGTGATCTGTCGAGCAGGATATTCGATGATCCAGCGGCC 62  
DB 3064 CGAGAGCCCAAGACAGAGTGTCCAGAGGAAGAAGAGAGTGAGGAGCGCGAG 3123  
QY 63 GCACAGCGCTGTGGGTAAAGGCTGTTGTAGAAAGCTGACCCAAACAG 122  
DB 3124 ACAGTGTCTCGGAAGGAAGAAGTTGCAGGAAGAGAGCGCGAGCGCAAGAGAG 3183  
QY 123 CTCCTGATGATCTTTTAAATGCGTCGTCATCTGCTCTGTAA 164  
DB 3184 CTCAGGAGGAAGAGAGACACTACGGAAGTGGAGCGCAA 3225

RESULT 13  
US-08-800-644-93  
Sequence 93, Application US/08800644  
Patent No. 5958752  
GENERAL INFORMATION:  
APPLICANT: Steiert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu

APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,644  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/056,200  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-800-644-93

Query Match 2.2% Score 32.4; DB 2; Length 9551;  
Best Local Similarity 50.0%; Pred. No. 6.7;  
Matches 81; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 3 CGGAACCTCCATGTGGCGGTGATCTGTCGAGCAGGATATTCGATGATCCAGCGGCC 62  
DB 3064 CGAGAGCCCAAGACAGAGTGTCCAGAGGAAGAAGAGAGTGAGGAGCGCGAG 3123  
QY 63 GCACAGCGCTGTGGGTAAAGGCTGTTGTAGAAAGCTGACCCAAACAG 122  
DB 3124 ACAGTGTCTCGGAAGGAAGAAGTTGCAGGAAGAGAGCGCGAGCGCAAGAGAG 3183  
QY 123 CTCCTGATGATCTTTTAAATGCGTCGTCATCTGCTCTGTAA 164  
DB 3184 CTCAGGAGGAAGAGAGACACTACGGAAGTGGAGCGCAA 3225

RESULT 14  
US-08-374-686-4/C  
Sequence 4, Application US/08374686  
Patent No. 5616474  
GENERAL INFORMATION:



GenCore version 5.1.6  
Copyright (c) 1995 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 6, 2003, 06:07:31 : Search time 273 Seconds

(without alignments)  
3245.237 Million cell updates/sec

Title: US-10-086-082-1

Perfect score: 1442

Sequence: 1 ccgggaactccatgtggc.....aatgcaattattgatacc 1442

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq-length: 6

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*

1: /cqn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cqn2\_6/ptodata/2/pubpna/PCI\_NEW\_PUB.seq.\*  
3: /cqn2\_6/ptodata/2/pubpna/US56\_NEW\_PUB.seq.\*  
4: /cqn2\_6/ptodata/2/pubpna/US36\_PUBCOMB.seq.\*  
5: /cqn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cqn2\_6/ptodata/2/pubpna/PCI05\_PUBCOMB.seq.\*  
7: /cqn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cqn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cqn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cqn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cqn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cqn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cqn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cqn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1442	100.0	1442	9	US-10-086-082-1
C 2	37.6	2.6	3902	9	Sequence 1, Appli
C 3	36.2	2.5	461	10	Sequence 88, Appl
C 4	35.8	2.5	941	9	Sequence 411, App
C 5	35.2	2.4	429	9	Sequence 484, Appl
C 6	35	2.4	724	9	Sequence 7, Appl
C 7	34.4	2.4	596	9	Sequence 60, Appl
C 8	34.4	2.4	596	9	Sequence 310, App
C 9	34.2	2.4	3188	9	Sequence 310, App
C 10	34	2.4	4153	9	Sequence 14, Appl
C 11	33.8	2.3	481	10	Sequence 123, App
C 12	33.8	2.3	1068	9	Sequence 372, App
C 13	33.8	2.3	1839	10	Sequence 3151, Ap
C 14	33.8	2.3	3309400	9	Sequence 6010, Ap
C 15	33.6	2.3	284	10	Sequence 1, Appli
C 16	33.2	2.3	415	10	Sequence 2052, Ap
C 17	33.2	2.3	671	9	Sequence 2327, Ap
C 18	33.2	2.3	671	9	Sequence 346, App
C 19	33.2	2.3	1049	9	Sequence 346, App
					Sequence 358, App

#### ALIGNMENTS

##### RESULT 1

US-10-086-082-1

Sequence 1, Application US/10086082

Publication No. US2003008742A1

GENERAL INFORMATION:

APPLICANT: Briceon, Walter

APPLICANT: Naughton, Andrew

APPLICANT: Robins, Karen

APPLICANT: Shaw, Nicholas

APPLICANT: Zimmermann, Thomas

TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)

FILE REFERENCE: 3213

CURRENT APPLICATION NUMBER: US/10/086.082

CURRENT FILING DATE: 2002-02-28

PRIOR APPLICATION NUMBER: US/09/214.679

PRIOR FILING DATE: 1999-12-30

NUMBER OF SEQ ID NOS: 14

SOFTWARE: FastSeq for Windows Version: 3.0

SEQ ID NO 1

LENGTH: 1442

TYPE: DNA

ORGANISM: Klebsiella oxytoca

US-10-086-082-1

Query Match

Best Local Similarity 100.0%; Score 1442; DB 9; Length 1442;

Matches 1442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCGGGAACCTCCATCGCGCGTATCTCGTCCAGCAGGATATTCGGATGATCCAGCGG 60

Db 1 CCGGGAACCTCCATCGCGCGTATCTCGTCCAGCAGGATATTCGGATGATCCAGCGG 60

Oy 61 CCGCAGCAGCGCTGCGGTAATGGATAAGGCGCTGTTAGAAAGCGTGACCCAAAC 120

Db 61 CCGCAGCAGCGCTGCGGTAATGGATAAGGCGCTGTTAGAAAGCGTGACCCAAAC 120

Oy 121 AGCTCTGATGATCTTTTAATGCGTCGTCATCTGCTCTGTAACCTAAACGCTATAAAT 180

Db 121 AGCTCTGATGATCTTTTAATGCGTCGTCATCTGCTCTGTAACCTAAACGCTATAAAT 180

Sequence 5, Appli  
Sequence 9, Appli  
Sequence 7, Appli  
Sequence 29598, A  
Sequence 13036, A  
Sequence 1257, Ap  
Sequence 7836, Ap  
Sequence 11, Appli  
Sequence 2, Appli  
Sequence 7, Appli  
Sequence 8, Appli  
Sequence 5, Appli  
Sequence 6, Appli  
Sequence 3, Appli  
Sequence 28, Appli  
Sequence 135, App  
Sequence 135, App  
Sequence 117, App  
Sequence 6205, App  
Sequence 1, Appli  
Sequence 22134, A  
Sequence 204, App  
Sequence 204, App  
Sequence 43, Appli  
Sequence 283, App  
Sequence 983, App

181	ACGTGGAGAAATACATATGAAATGGTTGGAGAAATCCATTATGGCCAAATCGCGGTTCTGG	243
181	ACGTGGAGAAATACATATGAAATGGTTGGAGAAATCCATTATGGCCAAATCGCGGTTCTGG	243
241	TGCGGGGCTAAACCGGTAAACGCATCACCTGACGGGAAGAAATGCAAAAGAGHTTCATTA	303
241	TGCGGGGCTAAACCGGTAAACGCATCACCTGACGGGAAGAAATGCAAAAGAGHTTCATTA	303
301	CACCATTTGGCGCTTATTCCACACCCGCTCTGACCATCGAAACCGGCTGACGGGATTAATCT	360
301	CACCATTTGGCGCTTATTCCACACCCGCTCTGACCATCGAAACCGGCTGACGGGATTAATCT	360
361	CGACACTCGAGATGCTTTTGAAGGTGCTATCAATTCGAAACAGAAATATTCCGAGTCAGTT	420
361	CGACACTCGAGATGCTTTTGAAGGTGCTATCAATTCGAAACAGAAATATTCCGAGTCAGTT	420
421	GCATAAAATGGGCTTTCTCAACCCACAAAGCGACCGATCATGGTCAATGGCGCGTAGAA	483
421	GCATAAAATGGGCTTTCTCAACCCACAAAGCGACCGATCATGGTCAATGGCGCGTAGAA	483
481	AGGTGATGTCGGCTGTCTATATCCAAATCCATGTTTGGCGGAGCGGCTTGATTCCTAAGC	540
481	AGGTGATGTCGGCTGTCTATATCCAAATCCATGTTTGGCGGAGCGGCTTGATTCCTAAGC	540
541	CATCTGCGCCATGANTCCGCAATTTTGGCGGACTGACCGGACCGAGCTGACGCCCATGCT	600
541	CATCTGCGCCATGANTCCGCAATTTTGGCGGACTGACCGGACCGAGCTGACGCCCATGCT	600
601	CAATGATCGCTGGCCAGAAAAGTGGCCATGATTAACTCGACAGCTGAAAAGGCTTACTG	660
601	CAATGATCGCTGGCCAGAAAAGTGGCCATGATTAACTCGACAGCTGAAAAGGCTTACTG	660
661	GAGCAAAACGCTATACGCTTCCGCTATAAACCCCATATTGSCACCTTCAGCGGTAATCGGAGA	720
661	GAGCAAAACGCTATACGCTTCCGCTATAAACCCCATATTGSCACCTTCAGCGGTAATCGGAGA	720
721	AATTGACTCAATCAATTCACCTGACCGCCAGACAAATCAGCGCGGAAATATGGATGTGCGGGA	780
721	AATTGACTCAATCAATTCACCTGACCGCCAGACAAATCAGCGCGGAAATATGGATGTGCGGGA	780
781	TATAGGACCAAGGAGTATTACCTATCTCGCGGTACGTGGCCCTGGAGGCGCGCTGTTTAT	840
781	TATAGGACCAAGGAGTATTACCTATCTCGCGGTACGTGGCCCTGGAGGCGCGCTGTTTAT	840
841	TGCTGATGCCATGCTTGTCAGGCTGATGGTGAGATTTCGGGACCGCATGACGATTTCG	900
841	TGCTGATGCCATGCTTGTCAGGCTGATGGTGAGATTTCGGGACCGCATGACGATTTCG	900
901	CTCAATCACCAACATCAAAAGTCGATTGTATCAAGAACTGGCAGCTTTCCTTGGCCACGAA	960
901	CTCAATCACCAACATCAAAAGTCGATTGTATCAAGAACTGGCAGCTTTCCTTGGCCACGAA	960
961	CGAGAAATGGGAAAATATATAGTATTGGCAGTGCAGTCCGCTGGAGATCGCAGCGG	1020
961	CGAGAAATGGGAAAATATATAGTATTGGCAGTGCAGTCCGCTGGAGATCGCAGCGG	1020
1021	AATTGCTATTCGGACTTAATTTACTGGCTGGTGAAGAGCTTTGGCTTCGAAACAATGGGA	1080
1021	AATTGCTATTCGGACTTAATTTACTGGCTGGTGAAGAGCTTTGGCTTCGAAACAATGGGA	1080
1081	TGCTACATGCTTCTCGATCATATGGCGCAAGTGGCGCTGGGCAACATGGTCTACGCCCAA	1140
1081	TGCTACATGCTTCTCGATCATATGGCGCAAGTGGCGCTGGGCAACATGGTCTACGCCCAA	1140
1141	ATACACCGTGGCGGATGCTGTAAGTATGGCAGTGCAGTCCGCTGGAGATCGCAGCGG	1200
1141	ATACACCGTGGCGGATGCTGTAAGTATGGCAGTGCAGTCCGCTGGAGATCGCAGCGG	1200
1201	TGAACATTACCGCGATGTATATCGGAGTATATGTTAGTTCGAAATATGGCTATTTTAA	1260
1201	TGAACATTACCGCGATGTATATCGGAGTATATGTTAGTTCGAAATATGGCTATTTTAA	1260
1261	CAGCTAAACGAGTGTATATGGCGGCAAGATACACCAATATATGGTATTTACTTACTCG	1320

DB	1261	CACCTAAGACAGGTGGCATATGGGGCCAGATACACCCA	CAATATTGGTTTACTCTTC	1320
QY	1321	TTTCAGCGGAGTACGGCGGCACAAAGAGTTGTGTCA	CAATGGCGGAGCAACCCAGGCTATT	1380
DB	1321	TTTCAGCGGAGTACGGCGGCACAAAGAGTTGTGTCA	CAATGGCGGAGCAACCCAGGCTATT	1380
QY	1381	GCCGAATAATCAAAATGCGGCATCAACGGCAGACCA	CTCAATGCAATTCATTGGAT	1440
DB	1381	GCCGAATAATCAAAATGCGGCATCAACGGCAGACCA	CTCAATGCAATTCATTGGAT	1440
QY	1441	CC 1442		
DB	1441	CC 1442		

RESULT 2  
 US-10-108-605-88/c  
 : Sequence 88, Application US/10108605  
 : Patent No. US20020160934A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Broadus, Julie  
 : APPLICANT: Stam, Lynn  
 : APPLICANT: Bachmann, Jane  
 : APPLICANT: Kandar, Kim  
 : TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT  
 : TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF  
 : FILE REFERENCE: 311348  
 : CURRENT APPLICATION NUMBER: US/10/109,605  
 : CURRENT FILING DATE: 2002-03-27  
 : PRIOR APPLICATION NUMBER: US 03/763,142  
 : PRIOR FILING DATE: 2001-01-16  
 : PRIOR APPLICATION NUMBER: US 62/176,418  
 : PRIOR FILING DATE: 2000-01-14  
 : NUMBER OF SEQ ID NOS: 361  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 88  
 : LENGTH: 3902  
 : TYPE: DNA  
 : ORGANISM: Drosophila melanogaster  
 : US-10-108-605-88

QY



APPLICANT: Yu, Yang  
APPLICANT: Rameka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Krickler, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Huban, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
TITLE OF INVENTION: Thaliana  
FILE REFERENCE: 2027 (PARA-015PRV)  
CURRENT APPLICATION NUMBER: US/09/770,444  
CURRENT FILING DATE: 2003-01-26  
PRIOR APPLICATION NUMBER: 60/178,502  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 411  
LENGTH: 461  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-444-411

Query Match 2.5% Score 36.2; DB 10; Length 461;  
Best Local Similarity 50.7%; Pred. No. 0.23;  
Matches 112; Conservative 0; Mismatches 108; Indels 1; Gaps 1;

QY 63 GCACAGCCGTGGCGTAATGGAATAAGGCGTGGTTGTAGAAAGCTGACCCACAA-CA 121  
DB 448 GAAGAGGACCGCTCTTTTATGATGATACATAGCTGCTGAGTATAGTCTTCACAGTANG 183  
QY 122 GCTCTCTCATGATCTTTTAATGCGTGGTCATCTGGCTGTGAAGCTAAGCGGTATAAATA 181  
DB 368 GAACCCGCTTCGCTGATACCTATCATCTCTTCTGTACAGAGCAATGGGAT 329  
QY 182 COTGGAGAAATACATATGAATGGTGGAGAAATCAATATGGGCAACGGGGCTTGGT 241  
DB 328 ACGGATGAGAAAGTTTGAATCTGTGCACCTTCTCATCAGGCCAAAGTTGGACAGAA 263  
QY 242 CGCGGCGCTAAGCGGTACGCACTCACTGACGAGAGAAAT 282  
DB 268 ATGGGTGTTTCTTTTATGACTCTTTTGA-TGACAGAAAT 228

RESULT 4  
US-10-123-155-464  
Sequence 464, Application US/10/23155  
Publication No. US2003068794A1

GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tamas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME  
FILE REFERENCE: P3330RIC30  
CURRENT APPLICATION NUMBER: US/10/123,155  
CURRENT FILING DATE: 2002-04-15  
Prior Application removed - See Palm. or File Wrapper  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 464  
LENGTH: 941  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-123-155-464

Query Match 2.5% Score 35.8; DB 9; Length 941;  
Best Local Similarity 6.2%; Pred. No. 0.51;  
Matches 47; Conservative 228; Mismatches 482; Indels 0; Gaps 0;  
QY 403 GGATATCCGAGCCAGTGTCTAAATAATGCCCTTTCTCAACCCCAAAACGCGACGATCAI 462  
DB 164 KSTYRTKEGRRLASTOFEPETAARNAFFCFDEPAFKASFIRPRHIALSNPLVK 223  
QY 463 GTCAATGGCGGAGAAAGTGATGTGCTGCTGTCTATATCAATCATCTGTCGCCCG 522  
DB 224 SVTVAEGLIEDHFQVTVKMSIYLVAFIISDFESVSKITKSGVKVSVYAVPDKINQADYAL 283  
QY 523 CGCGGTGATCCCTACGCACTCGCCCATGATCCGCATTTTGGCGGACTGACCGGAC 582  
DB 284 DAAVTLLSFYEDYFSPYPLPKQDLAAIPDFQSGAMENWGLTIYRESALLFDEKSSASS 343  
QY 583 CGACCTGACGGGCAAGCTCAATGATCGGCGGCGGCAAGAAAGTGGCATGATTAAGCTGGA 642  
DB 344 KLGITVTVAHFLAQRHGFNLVTNEMWDLNLEGEAFKMFESVSVTHPELKVGDYFFGK 403  
QY 643 CAGTGAAGAAAGTCTACGAGGAGCAACCCCATGCTTCCCTATATACCCCATATGGCAC 702  
DB 404 CFDAEVDALNSSEPVSTPVENPAQIREMPDDVYDKGACILNMLREYLSADAFKSGIVQ 463  
QY 703 CTTGAGGATTCGCCAAGATGACTCAATCAATTCATGACGCCAGACAAATCACCGCGG 762  
DB 464 YLOKHSYKNTKNEDLNDSMASICTDGVKMDGFCRSQHSISSSSHHQHGVDVKTMMNT 523  
QY 763 GAATATGATGTCGGGATATAGACAGGAGATATACCTATCTCGCGGTACGTCGCCGC 822  
DB 524 WILQGGPLITITVGRGNVHKOFHYMKSGDAPDGYLWHVPLTFTTSKSNVHRELLK 583  
QY 823 TGGAGGCGCGCTGTTATTGCTGATGCCCATGCTTGTGACGGTGTGGTGAGATTTCGG 882  
DB 584 TKIDVLIIPREVEWIKNVGMNGYIVHPEDRGWDSLTGLLGTHTAVSSNDRASLINNA 643  
QY 883 GAGCGCAGTAGAGTTTCCTCAATCACCACCATCAAGATCGATTTGATCAAGAACTGGCA 942  
DB 644 FOLVSIKLSIFERKALDLSLYLKHETEIMPVFOGLNECIPMYKLMFKRDMNEVETQFKAFI 703  
QY 943 GCTTCTCTGCCAGGAATGGAGATGCCGAATAATATATGAGTATTGGCAGCTCACGTC 1002  
DB 704 TRLRLDLIDKQWTDEGSEVSEQLRSELLLLACVHNYQPCVQRAEGYFRKWKESNGLSL 763  
QY 1003 CTTGGAGGATGGCAGCGCAATTCATATCGCGACTTAATTTACTGGCTGTACAGACTT 1062  
DB 764 PVDVTLAVFVAGAGSTEGWDFLYSKYQFSLSSSEKSOIEFALCTONKEKLQWLDESFK 823  
QY 1063 TGGCTTCCAAACATGGGATGCTACATGCTTCTGAGTCAATGGCGCAAGTGGCGGTGG 1122  
DB 824 GDKTKTOEFPOLITLIGRNPVGPYLAOFKRNKNKLVQKFLVQKFLSSSIAHVMVGTINQS 883  
QY 1123 CAACATGCTGACCCCAATACACCGTGTGGCGGATG 1159  
DB 884 IRTPLIEVKGFSSLKENGSQLRCVQQTITETIEFNIG 920

RESULT 5  
US-10-146-575-7/C  
Sequence 7, Application US/10146575  
Publication No. US20030059800A1

	Query Match	2.4%;	Score 35;	DB 9;	Length 724;
	Best Local Similarity	8.8%;	Prod. No. 2.81;		
	Matches 51;	Conservative 173;	Mismatches 357;	Indels 0;	Gaps 0;
QY	257	GTACCGCATCACCTGACGGAGAAAGCAAAAAGACTTCATTACACCAATGGGACTTAT	316		

```
QY 406 TATTCGGAGCGAGTGTCTTAAATATGCGCTTCTGACAGCCCAAAAGGATGATCATCGT 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 589 K.KT.SSB.YYW.WN.SWY.YD.TCDNCC.C.C.CDN.CD.DRBTAN.ANY.S.SNYEG. 530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 466 CAATGGCGGAGAAAGTGTGCTCGCTGTCTATATCAATCCCAAGTTCGCGCGGG 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 529 .CT..CBTBBB.A.....W.B.SCC.MT.SBASTSDSAAUKCA.TTACSTCSBSS 470
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 CGTGTATCCCTACCGCATCTCGCGCATGATTCGCGCACTTTTGGGGACT-GACCGGACCG 584
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 CSNATANTCSSAASS.SNABAS.CSSBASS.SNATAST.SSAASSHSNATASTCSSAASS 410
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 585 ACCTGACGGCCATCTCAATGATCGCGTCCAGAAAGGTCGGCATGATTAATCACTGACA 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 .SNATASCSAASS.SNATASTCSSBASS.SNATASTCSSAASS.SNATASTCSS.ASS 350
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 645 GTGAAAGGCTACTGGAGCAAGCCCAATAGCTTCCTATATAACCCCATAT-TTGGCAAC 703
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 .SNABAS.CSSBASS.SNATACTCSSAASSHSNATASTCSSAASS.SNATASTCSSAASS 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 704 TIGACGGTATGCCAGAAATTAATCAATCAATCTGACGCCAGACAAATACAGCGCGGG 763
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 .SNATASTCSSBASSHSNATASTCSSAASS.SNATACTCSNAAYS.SNATASTCSSAASS 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 764 AATATGATGTCCGGATATAGGACCGAGGATAT 798
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 229 .SNATASTYSSBASS.SNATASTCSSAASS.SNAT 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT B

```
US-10-184-634-310/C
; Sequence 310, Application US/10184634
; Publication No. US20030069694A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC217
; CURRENT APPLICATION NUMBER: US/10184,534
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-310
```

Query Match 2.4%; Score 34.4; DB 9; Length 596;  
Best Local Similarity 18.7%; Pred. No. 1.1;  
Matches 74; Conservative 112; Mismatches 207; Indels 2; Gaps 2;

```
QY 406 TATTCGGAGCGAGTGTCTTAAATATGCGCTTCTGACAGCCCAAAAGGATGATCATCGT 465
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 589 K.KT.SSB.YYW.WN.SWY.YD.TCDNCC.C.C.CDN.CD.DRBTAN.ANY.S.SNYEG. 530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 466 CAATGGCGGAGAAAGTGTGCTCGCTGTCTATATCAATCCCAAGTTCGCGCGGG 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 529 .CT..CBTBBB.A.....W.B.SCC.MT.SBASTSDSAAUKCA.TTACSTCSBSS 470
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 CGTGTATCCCTACCGCATCTCGCGCATGATTCGCGCACTTTTGGGGACT-GACCGGACCG 584
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 CSNATANTCSSAASS.SNABAS.CSSBASS.SNATAST.SSAASSHSNATASTCSSAASS 410
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 595 ACCTGACGGCCATGURCAATGATTCGCTGCCAGAAAAGGTGGCATGATTAATCACTGACA 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 .SNATASCSAASS.SNATACTCSSBASS.SNATACTCSSAASS.SNATACTCSS.ASS 350
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 645 GTGAAAGGCTACTGACAGCAAGCCCAATAGCTTCCTATATAACCCCATAT-TTGGCAAC 703
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 .SNABAS.CSSBASS.SNATACTCSSAASSHSNATASTCSSAASS.SNATACTCSSAASS 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 704 TIGACGGTATGCCAGAAATTAATCAATCAATCTGACGCCAGACAAATACAGCGCGGG 763
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 .SNATACTCSSBASSHSNATASTCSSAASS.SNATACTCSNAAYS.SNATACTCSSAASS 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 764 AATATGATGTCCGGATATAGGACCGAGGATAT 798
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 229 .SNATASTYSSBASS.SNATASTCSSAASS.SNAT 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Query Match 2.4%; Score 34.2; DB 9; Length 3188;  
Best Local Similarity 51.7%; Pred. No. 4;  
Matches 78; Conservative 0; Mismatches 73; Indels 3; Gaps 3;

```
QY 2 CGGGAAC:CCATGTGGCGGTGATCTGTGTCGACAGAGATTTGCGCATGATCGACGGGC 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2025 CGTGCAGTGTCTGTGCTCTGCTCTACAGACCTTCATGTGTGTCACGGGATTCAGGAGG 1370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 CGCACAGGCTGTGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1969 TGACAGGGCTGAGGATACCGCAICATGATGAAGAGAGGATGATGATGATGATGATGAT 1910
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 GCTCTCTGATGATCTTTTAAATGCGCTGCTCAT 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1909 GGTCTTGAAGAGGATCTTCTGATCATGAT 1879
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 10  
US-10-153-668-123  
; Sequence 123, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAWA, Kenya  
; TITLE OF INVENTION: STAT6 Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10153,668  
; CURRENT FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/253,172

```

: PRIOR FILING DATE: 2001-05-25
: PRIOR APPLICATION NUMBER: US 60/316,031
: PRIOR FILING DATE: 2001-08-31
: PRIOR APPLICATION NUMBER: US 60/428,423
: PRIOR FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: JP 2001-157643
: PRIOR FILING DATE: 2001-05-25
: PRIOR APPLICATION NUMBER: JP 2001-260681
: PRIOR FILING DATE: 2001-08-30
: PRIOR APPLICATION NUMBER: JP 2001-313175
: PRIOR FILING DATE: 2001-10-10
: NUMBER OF SEQ ID NOS: 488
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 123
: LENGTH: 4153
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (175)...(3024)
US-10-153-568-123

Query Match
Best Local Similarity 52.9%; Score 34; DB 9; Length 4153;
Matches 73; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 196 TATGAANTGGTGGAGAAATCCATATGGGCAACGGGTTGGTGGCGGGGTAAAC 255
Dn ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1074 TATTAAGTGTGGTGGAGAAATCCATATGGGCAACGGGTTGGTGGCGGGGTAAAC 133
QY 256 GGTAAGCATCATCTGAGGAGAAATGCAAAAGAGTTTCAATACAGCATGAGGCTTA 315
Dn || || || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1134 GTCAAGAGAGGACTGGCTGGAGAGCAATAGAGNACATCTGCTTACAGCACTACTAG 1193
QY 316 TTCCACACCGTCTGTAC 333
Dn ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1194 TTCCAGAGCAGCTGAC 1211

RESULT 11
US-09-833-381-372/c
: Sequence 372, Application US/09833381
: Patent No. US20020132050A1
: GENERAL INFORMATION:
: APPLICANT: Robison, Keith E.
: TITLE OF INVENTION: No. US20020132090A1: Nucleic Acid and Protein Homologs
: FILE REFERENCE: 5800-119
: CURRENT APPLICATION NUMBER: US/09/833,381
: PRIOR FILING DATE: 2001-04-11
: PRIOR APPLICATION NUMBER: 09/516,448
: PRIOR FILING DATE: 2000-02-29
: NUMBER OF SEQ ID NOS: 2050
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 372
: LENGTH: 481
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)...(481)
: OTHER INFORMATION: n = A,T,C or G
US-09-833-381-372

Query Match
Best Local Similarity 51.7%; Score 33.8; DB 10; Length 481;
Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 16 TGGCGGTGATCTGGTGGAGAGATATTGGATGATACAGCGGGGCGGACAGGGGCTG 75
Dn ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172 TGGCGGTATCTTTATCACACCATATATGGCGGGTGGAGAGATGGGATGGGATAC 119
QY 76 CGGTAGGTATTAAGCGCTCTTTAGAAATGTCAGGCGGCAACAGACGCTCTGATGATC 135
Dn ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Dn 112 TGGTGAATGATCAGGGGAGGCGGTGGAAAGGGTACCCTCGAACAGATTTTTCATGCAC 53
QY 136 TTATAATCGTCGTCATCTGCTCTGTAA 164
Dn | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
52 CGCAACATCCATTACACCGGTGGGTGTTA 24

RESULT 12
US-09-738-626-3151
: Sequence 3151, Application US/09738626
: Publication No. US20020197605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738,626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 3151
: LENGTH: 1068
: TYPE: DNA
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-3151

Query Match
Best Local Similarity 64.9%; Score 33.8; DB 9; Length 1068;
Matches 50; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 81: GGTAGTGGCGCTGAGGCGCGCTGTTATTTGCTATGCCCATGCTTCTCAGGCTGATGG 870
Dn ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 GGTCTGTGGGTGGAAGGCGGTTTACCTGGTGAAGCGGCTTCTCGGTGGGTTGG 451
QY 871 TCAGATTTGCGGACCG 987
Dn || |||| | | | | |
452 ATAGGTTGGTGGCGG 468

RESULT 13
US-09-815-242-6013
: Sequence 6310, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: PROKARYOTES
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
```

PRIOR APPLICATION NUMBER: 60/255,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/237,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,952  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6010  
LENGTH: 1839  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1839)  
US-09-815-242-6010

Query Match 2.3% Score 33.8; DB 10; Length 1839;  
Best Local Similarity 51.7%; Pred. No. 3.8;  
Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 16 TGCCCGGTATCTGGTGGACAGATATTCGATGATCCAGCGCGGCCACACCGCTGTG 75  
DB 626 TGGCGCTATCTTATACACACAGATATTCGATGATCCAGCGCGGCCACACCGCTGTG 685  
QY 76 CGGTAATGGAATAGGCTTGTGTAGAACGCTGACCCCAACACACAGCTCTCTGATGAT 135  
DB 686 TGGTATGATATCAGGCGGCGGCGGTGGAACGGGTACCGTCGACACAGATTTTCATGCA 745  
QY 136 TTTAATGCGTGTCTATCTGTCTGTCTGTAA 164  
DB 746 CGCAACATCTTACACCGCTGCGCTGTATA 774

RESULT 14  
US-09-738-626-1/C  
Sequence 1, Application 75/79738626  
Publication No. US200201976C5A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIYA, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OKAZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 1  
LENGTH: 3309400  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 2.3% Score 33.8; DB 9; Length 3309400;  
Best Local Similarity 64.9%; Pred. No. 3.3e+02;  
Matches 50; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
QY 811 GGTAGTGTGGCGCTGGAGCGCGCTGTTTATTTGTATGATGCCCAATGCTGTGACGGGTGAAG 870  
DB 3046738 GGTCCCTGTGGCTGGAGGTGGCGTTTACCTGGTGAISCGGCTTCTCGGCGGGGTGTGG 3046739  
QY 871 TCAGATTTCCGGGACCG 887  
DB 3046738 ATAGGTTTGTCTCGGCC 3046722  
RESULT 15  
US-09-983-965-2052  
Sequence 2052, Application US/09983965  
Patent No. US20020137160A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Ningbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
FILE REFERENCE: 37-21(10297)C  
CURRENT APPLICATION NUMBER: US/09/983,965  
CURRENT FILING DATE: 2001-10-26  
PRIOR APPLICATION NUMBER: US 09/465,231  
PRIOR FILING DATE: 1999-12-15  
PRIOR APPLICATION NUMBER: US 60/113,678  
PRIOR FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 5912  
SEQ ID NO 2052  
LENGTH: 284  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
OTHER INFORMATION: Clone ID: 28-LIB3057-005-Q1-K1-G7  
US-09-983-965-2052

Query Match 2.3% Score 33.6; DB 10; Length 284;  
Best Local Similarity 48.9%; Pred. No. 1.3;  
Matches 90; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
QY 29 GTTCGAGCAGATATTCGATGATCCAGCGCGGCCACACAGCGCTGTGCGGTATGATAA 88  
DB 4 GTTCGAGGCGTCTTCTCTGATCCGCCAACCGCGCGCGCTGTGCGGTATGATAACT 90  
QY 89 AGSCCTGGTGTAGAAAGCTGACCCCAACACAGCTCTGTGATGATCTTTTATCGGTG 148  
DB 92 ACTCTCTCTTTGTAAAGTGTGACCAAGGGAAGAAACACACAGGTTTATCGCTGC 150  
QY 149 TCATCTGGCTCTGTAACTAAAGCTTATAATATGCTGGAGAAATACATATGAATGGTGG 208  
DB 151 TCAGCTTGTAGAGACAGAAAGATTGTAATTTTTCATGGAAGATGAAAGTTGTCA 210  
QY 209 GAAG 212  
DB 211 GGAG 214  
Search completed: July 6, 2003, 07:56:52  
Job time : 286 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compuser Ltd.

CM nucleic - nucleic search, using sw model

Run on: July 6, 2003, 04:04:05 ; Search time 1972 Seconds  
(without alignments)  
11842.744 Million cell updates/sec

Title: US-10-086-082-1

Perfect score: 1442

Sequence: 1 cccgggaactccatgtggc.....aatgcaattcatgtgaaac 1442

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8697743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em\_estba:\*\*

2: em\_esthum:\*\*

3: em\_estlin:\*\*

4: em\_estm:\*\*

5: em\_estov:\*\*

6: em\_estpl:\*\*

7: em\_estro:\*\*

8: em\_estc:\*\*

9: gb\_est1:\*\*

10: gb\_est2:\*\*

11: gb\_est3:\*\*

12: gb\_est4:\*\*

13: gb\_est5:\*\*

14: gb\_est6:\*\*

15: em\_estfun:\*\*

16: em\_estom:\*\*

17: gb\_gss:\*\*

18: em\_gss\_hum:\*\*

19: em\_gss\_inv:\*\*

20: em\_gss\_pla:\*\*

21: em\_gss\_vit:\*\*

22: em\_gss\_fus:\*\*

23: em\_gss\_mam:\*\*

24: em\_gss\_mus:\*\*

25: em\_gss\_oth:\*\*

26: em\_gss\_pro:\*\*

27: em\_gss\_rtd:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.6	2.7	280	10	AW380638
C 2	37.8	2.5	551	13	BM615374
C 3	36.2	2.5	533	10	AV551507
4	36.2	2.5	678	13	BM582997
5	36.2	2.5	1101	17	CNS00370
C 6	36	2.5	495	10	AV629284

C 7	35.6	2.5	482	17	BM767373
C 8	35.6	2.5	735	17	AG117547
C 9	35.2	2.4	611	10	BE442656
C 10	35.2	2.4	772	17	BM709682
C 11	35	2.4	266	9	AA020733
C 12	35	2.4	421	14	H75593
C 13	35	2.4	441	9	AA019639
C 14	35	2.4	492	12	AA001557
C 15	35	2.4	505	12	BM834211
C 16	35	2.4	691	17	B57961
C 17	34.8	2.4	328	9	AG109318
C 18	34.8	2.4	356	17	AG610392
C 19	34.8	2.4	422	17	AQ555596
C 20	34.8	2.4	475	9	AI743670
C 21	34.8	2.4	477	9	AI146857
C 22	34.8	2.4	482	24	BQ247372
C 23	34.8	2.4	502	9	AI289776
C 24	34.8	2.4	524	12	BG052847
C 25	34.6	2.4	511	17	AQ482699
C 26	34.6	2.4	562	17	AQ478968
C 27	34.4	2.4	410	12	BF484542
C 28	34.4	2.4	474	10	BE496530
C 29	34.4	2.4	483	9	AA833827
C 30	34.4	2.4	491	14	H08687
C 31	34.4	2.4	623	17	B69197
C 32	34.4	2.4	641	13	BM392441
C 33	34.4	2.4	641	13	BM393654
C 34	34.4	2.4	1100	17	CNS00DYG
C 35	34.2	2.4	466	13	BM136789
C 36	34.2	2.4	531	13	B246954
C 37	34.2	2.4	656	13	B234037
C 38	34.2	2.4	661	10	BE483900
C 39	34.2	2.4	903	13	BM018322
C 40	34	2.4	347	14	R63461
C 41	34	2.4	515	12	BF252201
C 42	34	2.4	540	12	BF252018
C 43	34	2.4	565	17	AQ430850
C 44	34	2.4	669	12	BF251067
C 45	34	2.4	760	12	BG574169

## ALIGNMENTS

RESULT 1  
AW380638 280 bp mRNA linear EST 04-FEB-2000  
RC2-HT0275-011199-011-f11 HT0275 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
ATTNERS  
TITLE  
JOURNAL  
COMMENT

Human.  
Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RCGP <http://www.udwig.org.br/RCGP-ESTES>.  
Unpublished (1999)  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Fudwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@udwig.org.br](mailto:asimpson@udwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.udwig.org.br/scripts/gethtml2.pl?tl=RC2&t2=RC2-HT0275-011199-011-f11&t3=1999-11-01&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 37

BASE COUNT:





ORGANISM Chlamydomonas reinhardtii  
Eukaryota: Viridiplantae: Chlorophyta: Chlorophyceae: Volvocales:  
Chlamydomonadales: Chlamydomonadae: Chlamydomonas:  
REFERENCE 1 (bases 1 to 495)  
AUTHORS Asamizu, E., Miura, K., Kuchino, K., Inoue, Y., Fukuzawa, H., Ohya, K.,  
Nakanura, Y., and Tabata, S.  
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
adapted cells of Chlamydomonas reinhardtii  
JOURNAL DNA Res. 7 (5), 305-307 (2000)  
MEDLINE 20539644  
COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizukazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

FEATURES  
source  
1..495  
/organism="Chlamydomonas reinhardtii"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="LCLO5sh05\_r"  
/note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:  
XhoI; The cDNA library was constructed from cells cultured  
in a carbon stress acclimatized condition in which carbon  
dioxide concentration in the bubbling gas was changed from  
5% to 0.04%"  
BASE COUNT 92 a 143 c 155 g 89 t  
ORIGIN  
Query Match. 2.5%; Score 36; DB 10; Length 495;  
Best Local Similarity 45.9%; Pred. No. 7.2;  
Matches 123; Conservative 0; Mismatches 145; Indels 0; Gaps 0  
QY 318 CCACACCGCTCCGACCATGACACCGGTGACCGGATATTCGACACATCGACATCTT 477  
Db 325 CGGGGCACTCCAGGAACCGCGGGGCGGTGACCGCTTTGCTACCTACTTCGAGTCTTC 246  
QY 378 TTGAAGTGTGTATCAATTCGGAACAGGATATTCGAGCCAGTTCGCTAAAGATCGCTTC 437  
Db 265 TTGACCTTGACGCAATCACAACATCTCTGGCGGCGCGCTTCACGAATCTCAGCTTC 206  
QY 438 TCACCCACAAACGGACCGCATGCTGTAATCGCGGGGAGAAAGGTGATGCTGCTGCT 497  
Db 205 TCAGAGTTGATCTTGAAGTCTCCGCTGCTCTCGAAGTGAGGGGTGCTGCCAGGATG 146  
QY 498 TCATATATGATCATGTTCCCGCGCGGCTTGATCTGATCCATCGCATCTCGCCATGATC 557  
Db 145 AGCAGCTCGCGGGCGGCGGACCATCATGCTTTCGCGCCCGCTCTCAGCGCGGTTCCG 56  
QY 558 CGCATTTTGGCGGACTGACCGGGACCGA 585  
Db 85 CGCTGGATGCGCACGGGGATGAGACCGA 58

RESULT 7  
BH767373/c  
LOCUS BH767373 482 bp DNA linear GSS 20-MAR-2002  
DEFINITION BMBAC348F06SP6.PSU Brugia malayi Genomic Bac Library 3 Brugia  
malayi genomic, DNA sequence.  
ACCESSION BH767373  
VERSION BH767373.1 GI:19565137  
KEYWORDS GSS.  
SOURCE Brugia malayi.  
ORGANISM Brugia malayi.  
REFERENCE Eukaryota: Metazoa: Nematoda: Chromadorea: Spirurida: Filariidae:  
Onchocercidae: Brugia.  
1 (bases 1 to 482)  
AUTHORS Whitton, C., Daub, J., Ware, J., Quail, M., Hall, N., Barrell, H., Foster,  
J., Guillian, D., Slatko, B. and Blaxter, M.  
TITLE Genome survey sequences from the human parasitic nematode Brugia  
malayi  
JOURNAL Unpublished (2000)

COMMENT Contact: Blaxter, M.  
Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JT, UK  
Tel: +44 131 650 6760  
Fax: +44 131 670 5450  
Email: mark.blaxter@ed.ac.uk  
Sequenced from the Brugia malayi BAC library constructed by Claire  
Whitton and Dr Mike Quail. The sequence was generated by The  
Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in  
collaboration with Mark Blaxter, ICAPB, University of Edinburgh,  
Edinburgh, UK.  
Seq primer: SP6 (ATTAGTGACACTATAG)  
Class: BAC ends.  
Location/Qualifiers:  
1..482  
/organism="Brugia malayi"  
/strain="JRS"  
/db\_xref="taxon:6279"  
/clone.lib="Brugia malayi Genomic Bac Library 3"  
/sex="Mixed (male and female)"  
/library.type="whole.parasite"  
/dev\_stage="microfilaria (L1)"  
/note="Vector: pBACE3.6; Site\_1: BamH I; Brugia malayi  
genomic DNA was partially cleaved with Sau3A I and size  
fractionated. 7,392 clones were generated with mean insert  
size ~48 kbp. The library was constructed by Claire  
Whitton, Blaxter Nematode Genetics Lab, University of  
Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing  
Unit, The Sanger Centre, Cambridge, UK."  
BASE COUNT 168 a 89 c 45 g 178 t  
ORIGIN  
Query Match. 2.5%; Score 35.6; DB 17; Length 482;  
Best Local Similarity 54.2%; Pred. No. 9.5;  
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0  
QY 72 TGTGGGTAAATGATAAAGCGCTGTGTAGAAAGCGTGACCCACACAGCTCTCTCAT 131  
Db 286 TGTGGGTATTTGTTATTTCTCTATGACGACAGATGAAGAAAGTAAAGCTTTCAGTT 227  
QY 132 GATCTTTTATCGTCGTCATCTGCTCTGTAACCTAAAGCTATATAATAGTGGAGAT 191  
Db 226 GATGATTATATGACAGCAATGAATTTGATATGAATGAAGGAATAATTTTTCAGAAA 167  
QY 192 AACATATGAAA 202  
Db 166 AAAAAGGTAAA 156

RESULT 8  
AG117547  
LOCUS AG117547 735 bp DNA linear GSS 03-NOV-2001  
DEFINITION Pan troglodytes DNA, clone: PTB-125G18.F, genomic survey sequence.  
ACCESSION AG117547  
VERSION AG117547.1 GI:16738066  
KEYWORDS GSS.  
SOURCE Pan troglodytes male lymphoblast DNA, clone PTB-125G18.F.  
ORGANISM BAC Library clone: PTB-125G18.F.  
Pan troglodytes  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Pan.  
1  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Tetoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of Library PTB  
JOURNAL Unpublished  
2 (bases 1 to 735)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Tetoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asac Fujiyama, The Institute of Physical

and Chemical Research (RIKEN). Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0445, Japan  
 (E-mail: chimbes@gs.riken.go.jp, URL: http://hgp.gs.riken.go.jp/  
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 Clones are derived from the chimpanzee HAC library PB3. This BAC end  
 was generated during the RAD process and may have higher chance of  
 clone tracking errors.

## PRIMERS

Sequencing: -21M13

## LIBRARY

Vector : pXSI45

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1..735

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-125G18.F"

/sex="male"

/cell\_type="lymphoblast"

/clone\_lib="PTB Chimpanzee Male BAC Library"

188 a 186 c 150 g 198 t 3 others

## FEATURES

source

Query Match 2.5%; Score 35.6; DB 17; Length 735;  
 Best Local Similarity 58.5%; Pred. No. 13;  
 Matches 62; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1315 TACTCTTCAGCGGAGTGCACGAGGTTGCACAAATGGCGGCGGCAACCCAG 1374

DB 123 TCCTCTTCAGCGGAGGAAATGAGCTCATTCCTATTTGTCTGCAGCGGACACAG 182

QY 1375 GCTATTGCGGAAATTAACTCAAAATGGCGGATCAACGGCAGACAC 1420

DB 183 GTCTCGACCAATGAATTAATTCGCTGCCATTCGAATCAAGCAG 228

## RESULT 9

BE442656

LOCUS

WHE1101\_D03\_G052S wheat etiolated seedling root normalized cDNA

Library Trilicium aestivum cDNA clone WHE1101\_D03\_G05, mRNA

sequence.

BE442656

VERSION BE442656.1 GI:9442089

KEYWORDS

SOURCE bread wheat.

ORGANISM

Trilicium aestivum

Eukaryota; Viridiplantae; Strophophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae

; Triticeae; Triticum.

1 (bases 1 to 611)

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han

, P.S., Hsiao, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,

Rausch, C.J., Seaton, C.L., Tong, J.C., and Zhang, E.

The structure and function of the expressed portion of the wheat

genomes - Normalized root cDNA library

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105959773

Fax: 5105959818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: StrataGene SK primer.

## FEATURES

source

location/Qualifiers

1..611

/organism="Trilicium aestivum"

/cultivar="Chinese Spring"

/db\_xref="taxon:4565"

/clone="WHE1101\_D03\_G05"

/clone\_lib="Wheat etiolated seedling root normalized cDNA

library"

/tissue\_type="Root"

/dev\_stage="Five day old etiolated seedling"

/lab\_host="E. coli DH10B"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid

pBluescript SK; Site\_1: EcoRI; Site\_2: XhoI; Seeds were

surface-sterilized, germinated and grown aseptically in

the dark at room temperature on filter paper with water,

mystatin and cefotaxime in covered crystallization

dishes. Roots were harvested. The tissue, total RNA, and

poly(A) RNA were prepared, a cDNA library was made in the

TJ Close lab (Choi, Close, Fenton) at the University of

California, Riverside. The cDNA clones were in vivo

excised to give pBluescript phagemids before

normalization was carried out. The mass excision of

phagemid library and normalization were done in HT Nguyen

lab by D. Zhang at Texas Tech University. Normalization

protocol used was that of Soares. Plasmid DNA

preparations and DNA sequencing were performed in the OD

Anderson lab (all other authors)."

BASE COUNT 181 a 111 c 153 g 166 t

## ORIGIN

Query Match 2.4%; Score 35.2; DB 10; Length 611;

Best Local Similarity 51.2%; Pred. No. 15;

Matches 82; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 719 GAATTGACCTCAATTCACCTGACGCCGACACAACTCACGGGGAATGATGTCGC 778

DB 141 GCAATGAGCAGCAAGAACTATCTCTGAGAGAGAAATGGGGGAATTTGACATAAG 200

QY 779 GATAAGCAGCAGGAGTATTACCTATCTGCGGTACCTGCGCTGGAGCGCCCTGT 836

DB 201 AATCTAAGCAGAGCTTCCAAAGTTATCTACCAATTTGTGTGAAGGAGCAATCTGAGT 260

QY 839 ATTGGTATGCCCATCTTCTCAGGTCGATGTCGATTT 878

DB 261 ACTGGGAAATGCACTTTTCCCAAGGTSATGGTGAAGTCT 300

## RESULT 10

BE709682/c

LOCUS

BOMEL07IF BO\_2\_3\_KB Brassica oleracea genomic clone BOMEL07, DNA

sequence.

BE709682

VERSION BE709682.1 GI:18797369

KEYWORDS

SOURCE GSS.

ORGANISM

Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids 1; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 772)

Town, C.D., Van Aken, S., Unterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other GSSs: BOMEL07IF

Contact: Chris Town

TIGR Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1..772

/organism="Brassica oleracea"

/strain="TOL000DH3"

```

/db_xref="taxon:3712"
/clone="BOMEL07"
/clone_lib="BO_2_3_KB"
/Note="vector: PHOS1; Site_1: HstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT      175 a   165 c   181 g   251 t
ORIGIN
Query Match      2.4%; Score 35.2; DB 17; Length 772;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 592 GCGATGCTCAATCGCTCCGACGAAAGGTGGCGATGATTAACTCGACAGTGAAGA 65;
DB 636 GGACTTCGCTATAGAGCGGCTCATCAAGCAGAGTTCCTTAGAANAATGTAAGAA 577
QY 652 GGTCTACTGGAGCAACGCCATAGCGTTCCTATTAACCCCATATTGCGACCTTGAGCGT 71;
DB 576 CGTCTAAGTCGAAGAACCGGCAAGAGAGCGGCTAGAACGCGTCAACAGCGGAGCTATGAT 517
QY 712 ATCGCCGAAATTCGCTCAATTCATTCCTCAAGCCAGACATCAAGCGGGGAGATA 767
DB 516 TTCATAGAAATAGGCCCATAGCTGCTATGACGGTTTATCTTACTTCCGGAAGA 46;

RESULT 1:
AA020733/c 264 bp mRNA linear EST 30-JAN-1997
LOCUS ze63b10.s1 Soares retina N2b4HR Homo sapiens cDNA clone
DEFINITION IMAGE:363643 3' similar to WP:K02A2.3 CEG279] HOMETANIDE-SENSITIVE
NA-K-Cl COTRANSPORTER 7, mRNA sequence.
ACCESSION AA020733.1 GI:1484516
VERSION EST
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 268)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favell,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le.N., Le.N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan.F., Thierry-Mieg,J., Trevisakis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63103
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2350 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 171.
Location/Qualifiers
1..268
/organism="Homo sapiens"
/db_xref="GDB:1280346"
/db_xref="taxon:9606"
/clone="IMAGE:363643"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue="retina"
/dev_stage="55 year old"
/lab_host="DH10R (ampicillin resistant)"
/Note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st
strand cDNA was primed with a Not 1 : - cllq(dT) primer [5'
TGTACCACTGAAAGTGGAGCGCGCGCTTTTITTTTTTTTITTTT 3'1.
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick K. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo.
BASE COUNT      50 a   72 c   71 g   71 t   4 others
ORIGIN
Query Match      2.4%; Score 35; DB 9; Length 268;
Best Local Similarity 50.6%; Pred. No. 9.8;
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 535 CTACGGCATCTGCGCATGATTCGGCATTTTGGCGGANTGACGGGACGACCTGACGSC 594
DB 167 CAAGGAGACTTCTCAGCATCAAGCGGAGTGGGAGAACTTGAACCACTCCACAGTGGCGG 108
QY 595 CATGCTCAATGATCCGCTGCCAGAAAGGTGGCATGATTAACTCGACAGTGAAGGT 654
DB 107 CATGCACACTGCGTGGCTGACGAGGTCTGTAAGAAATCCCGGACGCGCAAGCT 48
QY 655 CTACTGGAGCAACGCGATACGCTTCTCTATAAACCCCATATTG 698
DB 47 TGTTTTGGCTCAACATGCTTGGGCTCCCGCACCGCAATGGTG 4

RESULT 12
H75593/c 421 bp mRNA linear EST 01-NOV-1995
LOCUS y005104.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:232927 5', mRNA sequence.
ACCESSION H75593.1 GI:1049521
VERSION EST
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 421)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favell,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le.N., Le.N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan.F., Thierry-Mieg,J., Trevisakis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63103
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert size: 1012
High quality sequence stops: 366
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1012 Std Error: 0.00
Seq primer: M13Rpl
High quality sequence stop: 366.
Location/Qualifiers
1..421
/organism="Homo sapiens"
/db_xref="GDB:3786104"
/db_xref="taxon:9606"
/clone="IMAGE:232927"
/clone_lib="Soares fetal liver spleen INFLS"
FEATURES
Source

```

```

/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="RH106 (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: p773D (Pharmacia) with a
with a modified polylinker; Site: 1; Pac 1; Site 2; Eco RI;
1st strand cDNA was primed with a Pac 1 - oligo(dT) primer
5' AACTGGAGATTAATTAAGATCTTATTTTITTTTITTTT 3';
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac 1 and cloned into the Pac 1
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      89 a 104 c 107 g 113 t      8 others
ORIGIN

```

```

Query Match      2.4%; Score 35; DB 14; Length 421;
Best Local Similarity 51.6%; Pred. No. 13;
Matches 80; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1044 ACTGGCTGTAGAACTTGGCTTCGACAAATGGGATCGCTACATGCTTCGAGTCAT 1103
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 242 ACTGGCAGAGGATGACTTGGCTTCGCGAGGAGATGATCTCAACACAGCTGACAAAT 183
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1104 GCGGCAAGTGGCGTGGCCACATGGTGGACCCCAATACACCTTGGCGGATGCGA 1163
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 182 CCACAGGGAGTCTTGGGCGGAATGGCGGTGCACAGTTACACAGTTGGGGTGAACC 123
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 1164 ACAAAACCTGTTAGTTTAGTAGGGAATAACTAAC 1198
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 122 ACTATACCTTTAATAGAGTGATCTTCTACCC 88
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 13
AA019639/c
LOCUS      441 bp mRNA linear EST 30-JAN-1997
DEFINITION z62q01.s1 Soares retina N2b4HR Homo sapiens cDNA clone
IMAGE:363600 3' similar to WP:K02A2.3 CE02791 BUMETANIDE-SENSITIVE
NA-K-Cl COTRANSPORTER 1, mRNA sequence.

```

```

ACCESSION  AA019639
VERSION     AA019639.1 GI:1453067
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 441)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,P.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gier,M., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Nardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
Underwood,K., Wohlmann,P., Worster,R., Wilson,R. and Marita,M.
Generation and analysis of 28C,000 human expressed sequence tags
Genome Res. 5 (9): 837-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through INL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2010 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 360.
Location/Qualifiers
1. 441
/organism="Homo sapiens"
/db_xref="GDB:1280303"
/ab_xref="taxon:9606"
/clone="IMAGE:363600"
/clone_lib="Soares retina N2b4HR"

```

```

FEATURES
Source

```

```

/sex="male"
/clone_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: eye; Vector: p773D (Pharmacia) with a
modified polylinker; Site: 1; Not 1; Site 2; Eco RI; 1st
strand cDNA was primed with a Not 1 - oligo(dT) primer [5'
1GTACCAATCTGAAGTGGAGGCGGCCCTTTTITTTTITTTTITTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not 1 and cloned into
the Not 1 and Eco RI sites of a modified p773 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)- RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Frederick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT      65 a 137 c 120 g 113 t      6 others
ORIGIN

```

```

Query Match      2.4%; Score 35; DB 9; Length 441;
Best Local Similarity 50.6%; Pred. No. 14;
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 535 CTACGGCATCGCCCAATGATTCGGCATTTTGGGGACTACCGGGACCCAGCTGACGGC 594
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 CAGGACTCTTTCAGCATGAAGCGGAGTGGGAGAACTTGAACCACTCCCAACGTGGCGC 108
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 595 CATGCTCAATGATCGCGTGCAGAAAAGTGGCGCATGATTAAACTTCACAGTGAAGAGGT 654
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 107 CATGCACACTGNCGTGGCGCTGAACGAGGTCACTCGTCAAGAAATCCGGGACCCCAAGCT 48
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 655 CTACTGGAGCAAGCCCATAGCTTCCCTATAAACCCCATATTC 698
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 47 TGTTCGTCACATGCTGCGCTCCCGCAACCCGCAATGCTG 4
Db      ||||| | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 14
AA001557/c
LOCUS      492 bp mRNA linear EST 29-NOV-1996
DEFINITION z646D06.s1 Soares retina N2b4HR Homo sapiens cDNA clone
IMAGE:362027 3' similar to WP:K02A2.3 CE02791 BUMETANIDE-SFNSITIVE
NA-K-Cl COTRANSPORTER ; contains Alu repetitive element., mRNA
sequence.

```

```

ACCESSION  AA001557
VERSION     AA001557.1 GI:1437022
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 492)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 3433 Std Error: 0.00
Seq primer: mob.REGA+PT
High quality sequence stop: 233.
Location/Qualifiers
1. 492
/organism="Homo sapiens"
/db_xref="GDB:1278730"

```

```

FEATURES
Source

```

source 1..505  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC lPIG"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: Not I; Site\_2: Sal I;  
Library made from pooled tissue from day 11, 13, 15, 20,  
and 30 embryos."  
BASE COUNT 104 a 121 c 158 g 122 t  
ORIGIN  
Query Match 2.4% Score 35; DB 12; Length 505;  
Best Local Similarity 55.3%; Pred. No. 15;  
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 2 CCGGGAACTCCATGTGGCGGTGATCTCTGGTCGAGCAGGATATTCGGATGATCCAGCGGC 61  
DB 334 CCTCAATCTTCACCTGGTCTCAGTCTGCTGCTGAAGGTGAATCGCGGTGCTTTGGGGAT 393  
QY 62 CGCACAGCGCTGTGGGTATGATGAAGGCTGGTTGTAGAAACGCTGACCCCAACA 121  
DB 394 TGAATACGTGGTCAGGACCGTGGCCCTGCAGGCAGAGGAGCTGACCCCTCCAGCA 453  
QY 122 GCT 124  
DB 454 GCT 456  
Search completed: July 6, 2003, 06:40:28  
Job time : 1978 secs

/db\_xref="taxon:9606"  
/clone="IMAGE:362027"  
/clone\_lib="Soares retina N2b4R"  
/sex="male"  
/tissue\_type="retina"  
/dev\_stage="55 year old"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Ordan; eye: Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; list  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCAGAGGAGCGCGCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p773 vector  
(Pharmacia). The retinas were obtained from a 55 year old  
Caucasian and total cellular poly(A)<sup>+</sup> RNA was extracted 6  
hrs after their removal. The retina RNA was kindly  
provided by Roderick R. McInnes M.D. Ph.D. from the  
University of Toronto. Library constructed by Bento  
Soares and M.Patima Bonaldo."  
BASE COUNT 80 a 141 c 131 g 129 t 1; others  
ORIGIN  
Query Match 2.4% Score 35; DB 9; Length 492;  
Best Local Similarity 50.6%; Pred. No. 15;  
Matches 83; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 535 CTACGGCATCTGGCCATCTTCGTCATTTTGGGAGTGTGACCGAGCGAGCGCG 594  
DB 167 CAAGACTTCTTCCACCAAGACCGGAGCTGGGAACTTGAACGAGTCCACGTGCGCG 108  
QY 595 CATGCTCAATGATCGGTGGCAGAAAGTCCGATGATTAAGTCTGACAGTGAAGAGCT 654  
DB 107 CATGCACACTGCTGGCGGTGACGAGGATCATGTTGAAGAAATCCCGGAGCGAAGCT 48  
QY 655 CTATGGAGCAACGCCAATAGCTTCTCTATTAACCCGATATG 699  
DB 47 TGTTTGTCAACATGCTGGCGCTCCCGCAAGCGCAATGGTG 4  
RESULT 15  
BG834211  
LOCUS 505 bp mRNA linear EST 25-MAY-2001  
DEFINITION 352345 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.  
ACCESSION BG834211  
VERSION BG834211.1 GI:14199231  
KEYWORDS EST.  
SOURCE Pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 505)  
AUTHORS Fabrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laedreid,W.W.  
and Keele,J.W.  
TITLE Design and use of two pooled tissue normalized cDNA libraries for  
EST discovery in swine  
JOURNAL Unpublished (2000)  
COMMENT Contact: Smith TPI  
USDA, ARS, US Meat Animal Research Center  
PO Box 156, Clay Center, NE 68933-0156, USA  
Tel: 402 762 4356  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 12  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACACAGTATGACCAT  
BACKWARD: GTTTCACAGTCACGAG  
Plate: 113 row: 1 column: 2  
Seq primer: ATTAGGTGACACTATAG.  
Location/Qualifiers

GenCore version: 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 15:36:56 : Search time 82 seconds  
(without alignments)  
533,002 Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1745  
Sequence: 1 MWLFESIMAKRGVAGCKP.....GNGVPRVYVGAMLNKLLV 324

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1980.DAT:
2:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1981.DAT:
3:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1982.DAT:
4:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1983.DAT:
5:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1984.DAT:
6:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1985.DAT:
7:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1986.DAT:
8:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1987.DAT:
9:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1988.DAT:
10:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1989.DAT:
11:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1990.DAT:
12:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1991.DAT:
13:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1992.DAT:
14:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1993.DAT:
15:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1994.DAT:
16:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1995.DAT:
17:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1996.DAT:
18:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1997.DAT:
19:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1998.DAT:
20:	/SID22/qcgdata/geneseq/geneseq-emb1/AA1999.DAT:
21:	/SID22/qcgdata/geneseq/geneseq-emb1/AA2000.DAT:
22:	/SID22/qcgdata/geneseq/geneseq-emb1/AA2001.DAT:
23:	/SID22/qcgdata/geneseq/geneseq-emb1/AA2002.DAT:

Pred. No. is the number of results predicted by change to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1731	99.0	328	19 AAW40263	K. oxytoca R-speci
2	355.5	20.3	3.5	21 AAB29631	Enterobacter cloac
3	233	13.3	296	23 ABB47924	Listeria monocytog
4	208	11.9	432	23 ABB93276	Herbicidally activ
5	198	11.3	452	21 AAG09207	Arabidopsis thalia
6	198	11.3	452	21 AAG50180	Arabidopsis thalia
7	178	10.2	404	21 AAG09208	Arabidopsis thalia
8	178	10.2	404	21 AAG50181	Arabidopsis thalia
9	177.5	10.2	575	19 AAW58856	C. acidivorans gam
10	174.5	10.0	400	21 AAG09209	Arabidopsis thalia

11	174.5	10.0	400	21 AAG50182	Arabidopsis thalia
12	148	8.5	33	19 AAW40275	K. oxytoca R-speci
13	130	7.4	307	23 ABB91277	Herbicidally activ
14	117	6.7	21	19 AAW40272	K. oxytoca R-speci
15	107	6.1	238	23 ABB49158	Listeria monocytog
16	105	6.0	842	23 ABB93909	Herbicidally activ
17	101	5.8	525	22 AAB70539	Human PR09 protein
18	101	5.8	525	22 AAB70540	Human PR010 protei
19	101	5.8	525	22 AAB70541	Human PR012 protei
20	101	5.8	526	22 AAB70542	Human PR012 protei
21	101	5.8	829	23 AAE15854	Human SE26 mature
22	101	5.8	853	23 AAE15853	Human SE26 protein
23	101	5.8	993	22 AAB70537	Human PR07 protein
24	101	5.8	994	22 AAB70538	Human PR08 protein
25	101	5.8	994	23 AAB70543	Human PR013 protei
26	100.5	5.7	261	22 AAW40273	K. oxytoca R-speci
27	100	5.7	23	19 AAW40273	K. oxytoca R-speci
28	99	5.7	20	19 AAW40264	K. oxytoca R-speci
29	98	5.6	18	13 AAW40270	K. oxytoca R-speci
30	94	5.4	19	15 AAW40271	K. oxytoca R-speci
31	93	5.3	682	21 AAY51947	A. sydowi all-SST
32	91.5	5.2	194	22 AAB704716	Propionibacterium
33	91	5.2	1131	22 ABB24897	Novel human diagn
34	90	5.1	2748	22 ABB58843	Drosophila melanog
35	89.5	5.1	635	22 AAU35532	Haemophilus influe
36	88.5	5.1	428	23 ABB90560	Chlamydia pneumoni
37	88.5	5.1	443	20 AAY34467	Porphyromonas ging
38	88.5	5.1	451	20 AAY34343	Porphyromonas ging
39	88	5.0	773	23 ABB55204	Lactococcus lactis
40	87	5.0	850	22 AAG91859	C glutamicum prote
41	86.5	4.9	428	20 AAY35396	Chlamydia pneumoni
42	86	4.9	388	22 ABB26871	Novel human diagn
43	86	4.9	1734	22 AAB20500	Full armyworm Bt t
44	85.5	4.9	925	22 AAG92740	C glutamicum prote
45	85	4.9	345	21 AAG47748	Arabidopsis thalia

#### ALIGNMENTS

RESULTS 1  
AAW40263  
ID AAW40263 standard: Protein; 328 AA.  
AC AAW40263  
ET 16-JUN-1998 (first entry)  
DE K. oxytoca R-specific amidohydrolase protein.  
KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;  
KW 3,5,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.  
OS Klebsiella oxytoca.  
PK W09801568-A2.  
XX 15-JAN-1996.  
XX 10-JUL-1997; 97WO-EP03670.  
XX 03-MAR-1997; 97CH-0000500.  
XX 10-JUL-1996; 96CH-0001723.  
XX (LANZ) LONZA AG.  
XX Brieden W, Naughton A, Robins K, Shaw N, Tinschert A;  
XX Zimmermann T;  
XX WPI; 1998-101063/09.  
XX N-PSDB; AAV10449.  
XX Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation

PT - by stereoselective hydrolysis of corresponding racemic amide using  
PI microorganism or derived enzyme, used as drug intermediate  
PS Claim 5; Page 34-35; 68pp; German.  
XX  
CC This sequence represents a R-specific amidohydrolase isolated from  
CC Klebsiella oxytoca strain PRS1 which allows the microorganism to utilize  
CC 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide as its sole nitrogen  
CC source. This amidohydrolase is used in a process for preparing  
CC (R)-isotoms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide which  
CC is cheaper than prior art optical resolution of the racemate using  
CC dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.  
XX  
SQ Sequence 328 AA:  
Query Match: 99.0%; Score 1731; DB 19; Length 328;  
Best Local Similarity 99.4%; Pred. No. 7e-170;  
Matches 326; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
CY 1 MKWLEESIMAKRGVAGKRPVTHHTEEMOKERHYTIGPSTPVLTIPEGDRIIIVTRDA 60  
DB 1 MKWLEESIMAKRGVAGKRPVTHHTEEMOKERHYTIGPSTPVLTIPEGDRIIIVTRDA 60  
CY 61 FEGAINSEQDIPSOQLKMPFQNGPIVNGAKGQDLAVYIESMLPRGVDPYGCAMI 120  
DB 61 FEGAINSEQDIPSOQLKMPFQNGPIVNGAKGQDLAVYIESMLPRGVDPYGCAMI 120  
CY 121 PHFGGLTGDTLAMLNDPLPEKVMKIDSEKVKYKRIIPYKPHIGTUSVSPEDISIN 180  
DB 121 PHFGGLTGDTLAMLNDPLPEKVMKIDSEKVKYKRIIPYKPHIGTUSVSPEDISIN 180  
CY 181 SLTPDNHSGNMVDPDIGFGSITVLPVRAPGGRFLFISDAHACGGDGE:CGTAVFEFASITTI 240  
DB 181 SLTPDNHSGNMVDPDIGFGSITVLPVRAPGGRFLFISDAHACGGDGE:CGTAVFEFASITTI 240  
CY 241 KVDLIKXWOLSWPRMENAENIMS:IGSARPLEDATRIAYRDLIYWLVEDFGFEQNDAYML 300  
DB 241 KVDLIKXWOLSWPRMENAENIMS:IGSARPLEDATRIAYRDLIYWLVEDFGFEQNDAYML 300  
CY 301 SQCGKVRGLGNMVDPKYTVGAMLNKNLLV 328  
DB 301 SQCGKVRGLGNMVDPKYTVGAMLNKNLLV 328  
RESULTS  
AAB29631  
ID AAB29631 standard; protein; 315 AA.  
AC AAB29631;  
ET 23-FEB-2001 (first entry)  
XX Enterobacter cloacae stereoselective amidase.  
DE  
XX  
KW Amidase: stereoselective hydrolysis; alpha-amino acid amide;  
KW alpha-hydroxy acid amide; optically active product;  
KW recombinant production.  
XX  
OS Enterobacter cloacae.  
XX  
PN W0200063354-A1.  
XX  
PC 26-OCT-2000.  
XX  
PF 17-APR-2000; 2000WO-JP02492.  
XX  
XX 16-APR-1999; 99JP-0109329.  
XX  
PA (MITR ) MITSUBISHI RAYON CO LTD.  
XX  
PI Nakamura T, Yu F;  
XX  
XX WPI; 2000-672731/65.  
DR

DR N-PSDB; AAC64148, AAC64149.  
XX Novel amidase gene encoding protein: which stereoselectively hydrolyzes  
PT alpha-amino-acid amides and alpha-hydroxy-acid amides, with enhanced  
PT catalytic capability to yield optically active alpha-amino-acids and  
PI alpha-hydroxy-acids  
XX  
PS Claim 1; Page 20-22; 32pp; Japanese.  
XX  
CC The invention relates to a novel amidase, and the gene encoding the  
CC novel amidase, from Enterobacter cloacae. The novel amidase  
CC stereoselectively hydrolyses alpha-amino acid amides and alpha-  
CC hydroxy acid amides to give optically active alpha-amino acids and  
CC alpha-hydroxy acids. The invention also relates to variants of the novel  
CC amidase which retain activity, recombinant vectors and host cells  
CC containing the novel amidase gene, and recombinant production of the  
CC novel amidase. Genetically manipulated microorganisms comprising the  
CC novel amidase gene have remarkably elevated catalytic capability as  
CC compared to those produced by conventional methods. The present sequence  
CC represents the novel Enterobacter cloacae amidase.  
XX  
SQ Sequence 315 AA:  
Query Match 20.3%; Score 355.5; DB 21; Length 315;  
Best Local Similarity 31.9%; Pred. No. 6.5e-28;  
Matches 100; Conservative 42; Mismatches 136; Indels 35; Gaps 9;  
CY 23 RHUTEEMOKERHYTIGPSTPVLTIPEGDRIIIVTRDAFEGAIN---SEQDIPSOQLKMP 79  
DB 1 RHUTEEMOKERHYTIGPSTPVLTIPEGDRIIIVTRDAFEGAIN---SEQDIPSOQLKMP 79  
CY 80 F-LNPQNGPIVNGAKGQDLAVYIESMLPRGVDPYGCAMI:PHFGSLTGDTLAMLND 137  
DB 80 F-LNPQNGPIVNGAKGQDLAVYIESMLPRGVDPYGCAMI:PHFGSLTGDTLAMLND 137  
CY 61 FDKLNPVSGPVVVEGAQPGDALKVTLESFRPSG---FGWTANIPGFGLLAD-----QFSD 112  
DB 61 FDKLNPVSGPVVVEGAQPGDALKVTLESFRPSG---FGWTANIPGFGLLAD-----QFSD 112  
CY 138 PLPEKVMKIDSEKVKYKRIIPYKPHIGTUSVSPEDISINSLTPDNH-GGNMVDPI 196  
DB 138 PLPEKVMKIDSEKVKYKRIIPYKPHIGTUSVSPEDISINSLTPDNH-GGNMVDPI 196  
CY 113 PALTUQYDROGLTPCAFGRYGRVPLKPFAGTLGVAPAAAGHHVPPRRVGGNLDIIDL 172  
DB 113 PALTUQYDROGLTPCAFGRYGRVPLKPFAGTLGVAPAAAGHHVPPRRVGGNLDIIDL 172  
CY 197 GPGSITVPLVRAPGGRFLFISDAHACGGDGE:ICGTAVEFASITTIKVDLIKXWOLSWPRME 256  
DB 197 GPGSITVPLVRAPGGRFLFISDAHACGGDGE:ICGTAVEFASITTIKVDLIKXWOLSWPRME 256  
CY 173 AAGCTLWLPEVEGALFSGDTHAAGDGVGCGTALESAMCVVVKLVKMDPLKTPFA 232  
DB 173 AAGCTLWLPEVEGALFSGDTHAAGDGVGCGTALESAMCVVVKLVKMDPLKTPFA 232  
CY 257 NAENIMS-----IGSARPLEDATRIAYRDLIYWLVEDFGFEQNDAYMLSQCGKV 306  
DB 257 NAENIMS-----IGSARPLEDATRIAYRDLIYWLVEDFGFEQNDAYMLSQCGKV 306  
CY 307 RLGNMVD-PKYTV 318  
DB 307 RLGNMVD-PKYTV 318  
CY 293 RTSEIVDRNMVV 305  
DB 293 RTSEIVDRNMVV 305  
RESULTS  
AAB47924  
ID AAB47924 standard; Protein; 296 AA.  
XX AAB47924;  
XX  
XX  
ET 05-FEB-2002 (first entry)  
XX  
DE Listeria monocytogenes protein #625.  
XX  
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.  
XX  
XX Listeria monocytogenes.  
XX  
XX W0200177335-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-FR01118.  
XX

```

PR 11-APR-2000; 2000FR-0004623.
XX (INSP ) INST PASTER.
PA
XX Buchrieser C, Franquell E, Courze E, Russek C, Tsibi H, Dehoux P;
XX Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cosset B;
XX Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Roland JA;
XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
XX Chakraborty T, Doonan E, Hain T, Berche P, Charbit A, Durant L;
XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
XX MacKenzie E, De Pablos B, Wehland J, Kanst U, Estlin K, Hauf J;
XX Rose M, Voss H;
XX WPI: 2002-010914/01.
DR
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides
XX
XX Claim 6; SEQ ID NO 629; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
XX monocytogenes EGD-e (see ARA03041). The genome sequence and fragments of
XX it are useful for selecting probes and primers for detecting genes in L.
XX monocytogenes and related organisms, and for studying genetic
XX polymorphisms and other genomes. The present sequence is a protein
XX encoded by the genome sequence of the present invention. Proteins
XX expressed from the genome sequence are useful for raising specific
XX antibodies, identification of L. monocytogenes and related organisms, and
XX for biosynthesis and biodegradation, especially biosynthesis of vitamin
XX B12. The genome sequence and proteins encoded by it are also useful for
XX selecting compounds that regulate gene expression and cell replication
XX and modulate L. monocytogenes-related diseases. In addition, the genome
XX sequence and proteins encoded by it are useful in pharmaceutical and
XX vaccines compositions for the treatment or prevention of infections by L.
XX monocytogenes and related organisms.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 296 AA;
SQ
Query Match 13.3%; Score 233; DB 23; Length 296;
Best Local Similarity 25.0%; Pred. No. 2,6e-5;
Matches 74; Conservative 62; Mismatches 116; Indels 24; Gaps 2;
QY 28 ENKKEHYTIGYSTPVLTIEPGDRIIVD:PRAFEGAINSECHIPSOLLKMPFLNPGSP 82
DB 13 EMEKS-----TEPAIRVXGVSVAIKTKHENGCIHAKQLHNGEJMKQF-SPTIP 63
QY 88 IYVNGAKGRVLAVYIESMLPRGVDPYGCAMIPHFGLTGTCTITAMNDPFEKVMYK 147
DB 64 IYIEARFGDLLATIEKIELLGTVEFLNG--PNT-GITDRLITS-----NSTCYK 113
QY 148 LDSKAVVSKRHTLPYKPHGTLSVSPEDSIN-SLTPNKHGSMVDPDIPGASITVEL 256
DB 114 VENQNIYSEDIPIPIKTKGLL-----KTEELAFSKVPTKNGGLDPSKLTREGATIPY 169
QY 207 RAPGRRLFGDAACDGGEGTGAFAFASITKVDLIKNNWLSMRMENAKNINMSIGS 266
DB 170 EKYGASLVGNVRATTFGKITATSAAPAEVIRLQILANKRATPTTIHYHNEGLIAS 259
QY 267 ARPLEDATRIAYROLIYWLVEDFGEQWDAYMILASQCKVRLGNMVSPKYTVGAML 322
DB 230 DITIEKATOKTMHNWITLLIESDKWITFDALFLISLQADFQVCKLCKPNITSIKL 285
RESULT 4
ID ABB93276
XX ABB93276 standard; Protein; 432 AA.
XX
XX AC ABB93276;
XX

```

---

```

DT 31-MAY-2002 (first entry)
XX
XX Herbicidally active polypeptide SEQ ID NO 2487.
XX
XX Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
XX
XX W020210210-A2.
XX
XX 07-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-090892.
XX
XX 25-AUG-2001; 2001WO-090892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidner M;
XX
XX WPI: 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 2487; 261pp + sequence listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 432 AA;
SQ
Query Match 11.9%; Score 208; DB 23; Length 432;
Best Local Similarity 24.1%; Pred. No. 1.8e-12;
Matches 92; Conservative 44; Mismatches 140; Indels 106; Gaps 16;
QY 28 ENKKEHYTIGYSTPVLTIEPGDRIIVD:PRAFEGAINSECHIPSOLLKMPFLNPGSP 84
DB 20 QDQQLHKKHREIPVPAEVKAGEFFRVEIDAMGGVINKDNDSDIKN--LVLTTHHL 77
QY 85 NGPIMV-----NSAEKGVVLAVYIESMLPRGVDPYGCAMIPHFGLTGTCTITAMNDPL 139
DB 79 SGPIRVVDEEGVAARAGDLLAVEICNLGLPDEMGFTGSPDRENG-----GGFLTDHF 131
QY 140 PEKVMIKLDESKVYVSKHTLPYK-----HIGTILSVSP----- 174
DB 132 P-----CATKAIWYFEGIYAYSPOIPGVRFPGTHPGVIGTAPSNELLRIWNRERQ 183
QY 175 -EIDSTNSL-----TPD-----NHGSMVDPDIPGCSITYPLVRAPG 210
DB 184 LEESGVESLTCEIEGCTPEWERANEARTINGRENGCNDIKWSEGSKIYLPVPEG 243
QY 211 GRLFTGDHACAGDGEI--CGTAVFASITTKIVDLIKNNWLSM-----PRMENAENMS 263
DB 244 ANLSTGDMHFGSGDGEISFCG-AIEMSGFLELKCEIIRNGHMOEVLTPMGPTPLHVNPIFE 302
QY 264 IGSARP-----LEDAIRIAYROLIYWLVE---DFGEQWDAYMIL 300
DB 303 IGPVEPRFSEMLVFEGISVDESCKOHLDAT-VAYKRAVLNAIDYLFKEGYSKEOVYILL 361
QY 301 SCCG-KVRLGNMVSPKYTVGAM 321
DB 362 SCCPCEGRLSGIIVDSFNAVATL 383

```



RESULT 5  
AAG09207  
ID AAG09207 standard; Protein: 452 AA.  
XX AC AAG09207;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 7049.  
XX DE Arabidopsis thaliana.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP:033405-A2.  
XX PD C6-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-030:439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131149.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 05-JUL-1999; 99US-0142055.  
PR 05-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.



PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158212.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.

```
PR 29-OCT-1999; 99US-0132142.
Query Match: 11.3%; Score 198; DB 2; Length 452;
Best Local Similarity 24.9%; Pred. No. 2e-11;
Matches 100; Conservative 46; Mismatches 11; Indels 124; Gaps 19;

QY 28 EMQKEEHTGPGYTPVLTTPGDKIIVETDFAEGAI---NSGGLPSGLAKHFTNPQ 84
D 20 QQDQPLHNRWHPPEPPVAFVKGAEFFRVMIDAMGVKNDASDCKN-LVLTITHEL 77
QY 85 NGRINW-----NGAKGVLAIVYIESMLPROVDFGI-----CA---118
D 78 SGPIRVDEEGVAAGAGDELCAVEICNLAGPLGDEWGETGSGDFRNGGFTHEFTATKA 127
QY 119 -----MAP--HEGILT--GSDLTAMLNQPLFEKVRMKIDSEKVVYVSKRH---154
D 138 IWFEGIVAYSPQGVRFPGLTGGVIGTA---PSNEILLI-----WDFRQGL 144
QY 160 -----TLPKPHIGLTVSPE-----IDTSNLTPO-----NIGGRK 171
D 195 EESVESITLCEVVHQRPLACLPTRKGLGNIETEGTPEWERIANEAARTIPGEKGN 244
QY 192 DVDPGPGSITYPLVRAPGGRLFIQDAACQGGEL--CGFAVEFASLTITIKVLIKNC 249
D 245 DIKNLSRGSKYLVFVVEGANLSTGDMHFSQGGESLFCG-AIEMSGFLELKEILNGM 303
QY 250 LSW-----PRMENAIMSIGSARP-----LEDAITRIAYRDLIYW 284
D 304 QEYLTPMGPTPLHVNPIEIGFVPEFSEWLVFEGISVDFSGKHVLDAT-VAYKNAVLN 362
QY 285 LVE---DREGSDWAYMLSCG-KVLEGNMVDPKYTVGM 321
D 363 AIDYLFKFGYSKEOVYLLSCCPCEGRZSGIVDSPNAVATI 403

RESULT 7
AAG09208
ID AAG09208 standard; Protein; 404 AA.
AC AAG09208;
XX
XX
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7050.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
PN
XX
XX EP1033405-A2.
PD
XX
XX 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301419.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0121548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126783.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 04-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138546.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140591.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141542.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144066.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
```



XX 25-FEB-1999; 9905-0121825.  
PR 05-MAR-1999; 9905-0123180.  
PR 09-MAR-1999; 9905-0123548.  
PR 23-MAR-1999; 9905-0125788.  
PR 25-MAR-1999; 9905-0126264.  
PR 29-MAR-1999; 9905-0126785.  
PR 01-APR-1999; 9905-0127462.  
PR 06-APR-1999; 9905-0128234.  
PR 08-APR-1999; 9905-0128714.  
PR 16-APR-1999; 9905-0129845.  
PR 19-APR-1999; 9905-0130077.  
PR 21-APR-1999; 9905-0130445.  
PR 23-APR-1999; 9905-0130510.  
PR 28-APR-1999; 9905-0130891.  
PR 30-APR-1999; 9905-0131449.  
PR 30-APR-1999; 9905-0132048.  
PR 30-APR-1999; 9905-0132407.  
PR 04-MAY-1999; 9905-0132484.  
PR 05-MAY-1999; 9905-0132485.  
PR 06-MAY-1999; 9905-0132486.  
PR 07-MAY-1999; 9905-0132487.  
PR 11-MAY-1999; 9905-0132863.  
PR 14-MAY-1999; 9905-0134256.  
PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134259.  
PR 14-MAY-1999; 9905-0134221.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-0134769.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135134.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136392.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 16-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139809.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140695.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144894.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145087.  
PR 23-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 23-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 05-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 11-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149175.  
PR 16-AUG-1999; 9905-0149175.  
PR 16-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 27-AUG-1999; 9905-0151303.  
PR 27-AUG-1999; 9905-0151348.  
PR 27-AUG-1999; 9905-0151930.  
PR 01-SEP-1999; 9905-0152363.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 15-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156456.

PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159305.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161362.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 10.2% Score 178; DB 21; Length 404;  
Best Local Similarity 25.4%; Pred. No. 2e-09;  
Matches 94; Conservative 40; Mismatches 112; Indels 124; Gaps 19;  
QY 59 DAFEGAI--NSEDIPSGAKKPFNPNQPIW-----NGAEKGVLAAYIESMLPRG 110  
DB 3 DAMGGVTKDMSADIKN--LVLTTHHLSPIRVYDEERVAKAGGGLAIVEICNIGPFP 60  
QY 111 VDPYGI-----CA-----MIP--RFGILT--GGLTA 133  
DB 61 GDWGTGTSDFRNGSGHLDHFFPCA-KA-WYFEGIVAYSPQIFGVYRFDLTHGVZGA 120  
QY 134 MLDPLPEKVRMKDSEKIVYKSRH-----LTPYKPIGTLVSPE-----ID 177  
DB 121 -----PSNELLRI-----WDRERQLESCVESLTCEVYVQRPLACPTTKGGLIG 167  
QY 178 SINSLIPD-----NHGNNMVPITGPGSLIYTPVVRAPGGRLFGDAHACQ 222  
DB 168 NIEGCTEWSERIANEAARTIPRGNGCNDIKNLSRGSKIYLPVFEVGANLS-GMGFSQ 227  
QY 223 GIGGTI--CGTAVERASITIKVDLIKNNQLSW-----PRMENENINSIGSARF----- 259  
DB 228 GDGEISFCG-AIEMSGFLELKCEIIRNGOEYLTPEKPTPLHVNPIFEIGPVEPRFSEKL 286  
QY 270 -----LEDATSTAYRDLTYWIVE---DFGFEQWADYMLSCCG-KVRLSNW 311  
DB 287 VFEISVDESKOYLDAT-VAYKRAVLNAIDYLFKFGYSKEOVYLLSCCPGEGRLSGT 345  
QY 312 VDPKTYTGAM 321  
DB 346 VDSPNAVATL 355

RESULT 9  
AAW58856  
ID AAW58856 standard; Protein: 575 AA.  
XX

AC AAW58856;  
XX 23-JUL-1998 (first entry)  
DI C. acidivorans gamma-lactamase protein.  
XX Gamma-lactamase; enantiomer; bicyclic lactam; temperature stable;  
XX 2-azabicyclo(2.2.1)hept-5-en-3-one; therapeutic agent; antiviral;  
KW cardiac vasodilator.  
XX Comamonas acidivorans.  
XX WC9810075-A1.  
XX 12-MAR-1998.  
XX 01-SEP-1997; 97WO-GB02344.  
XX 03-SEP-1996; 96GB-0018340.  
XX (CHIR-) CHIROSCIENCE LTD.  
XX Brown KC, Lee CS, Wisdom RA;  
XX WPI: 1998-193625/17.  
XX N-ESDR: AAV11459.  
XX Lactamase enzyme specifically hydrolyzing (+)-enantiomer of 2-  
XX azabicyclo(2.2.1)hept-5-en-3-one - isolated from Comamonas  
XX acidivorans or prepared by recombinant DNA technology. Used as, e.g.  
XX cardiac vasodilator  
XX Claim 5; Page 17-20; 28pp; English.  
XX This sequence represents a gamma-lactamase capable of hydrolyzing an  
XX enantiomer of the bicyclic lactam 2-azabicyclo(2.2.1)hept-5-en-3-one.  
XX This enzyme has at least 1 of the following features: greater than 85%  
XX retention of activity after incubating at 40 deg. C for 4 hours, or  
XX >30% activity after incubating at 60 deg. C for 4 hours, hydrolysis at  
XX an initial concentration of 100 g racemic lactam plus 300 ml buffer and  
XX proceeding to at least 30 (especially at least 98%) hydrolysis of the  
XX (+) lactam with <5% hydrolysis of the (-) lactam. The enzyme is much  
XX more temperature-stable than previously identified (-)-gamma-lactamases  
XX used for the same purpose. It also enables the bioresolution to be  
XX carried out at very high substrate/product concentrations. This  
XX enzyme has applications as a therapeutic agent e.g. as an antiviral  
XX agent or as a cardiac vasodilator.  
XX Sequence 575 AA;  
QY Query Match 10.2% Score 177.5; DB 19; Length 575;  
Best Local Similarity 24.0%; Pred. No. 3.8e-09;  
Matches 69; Conservative 33; Mismatches 85; Indels 101; Gaps 12;  
QY 39 PYSTP-----VLITPQDRILIVDTDAFEGAI---NSEQDIPS-QLKMKPLN 82  
DB 44 PYDNPOVNRHHPDIPMAVWVEPGAFFKLEYDTMTCGA:KNDSAEVDYDLSTVHFL- 102  
QY 83 PONGPIWNGAEKGVLAAY-----YIESMLPRGV----D 112  
DB 103 --SGPVGVGAQPGDGLLVVDLLDLCARDOSLWFGNGFFSKONGGGLDEHFLAKSIND 160  
QY 113 PYGICAMIPH-----FGLTGTDLTAMI.NDPLPEKVRMKIKLDSKVKVYKSRHTLPYKPHI 167  
DB 161 PHGMFTKSRHIPGVNFAGLIHPGLIGCLPDP-----KML-----ASNRETI----- 202  
QY 168 GTLSVSPE-----IDSINSLIPDNHGGNMVDVDPGPG 199  
DB 203 GLIATDPPDRIPLGANPPNATTAHMQOMOGAEARDAKAAEGARTVPPRPHGNCNDKLSRG 262  
QY 200 SITYPVLRAPGRLFGIDAHACQGDGEIC-GTAVEFASITTIKVDLIK 246  
DB 263 SRVFFPVYVDGAGLSVGLHFSQGDGEITFWGPIEMPGVWHKVSLLIK 310

RESULT 10  
AAG09209  
ID AAG09209 standard; Protein: 400 AA.  
XX  
AC AAG09209;  
XX  
XX  
D- 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 7051.  
XX  
KW Protein identification; signal transduction; pathway: metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
DN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX  
PF 25-FEB-2000; Z000EP-C301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123348.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130610.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134368.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139761.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140931.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 32-AUG-1999; 99US-0146386.  
PR 32-AUG-1999; 99US-0146388.  
PR 32-AUG-1999; 99US-0146389.  
PR 33-AUG-1999; 99US-0147038.  
PR 34-AUG-1999; 99US-0147204.  
PR 34-AUG-1999; 99US-0147302.  
PR 35-AUG-1999; 99US-0147192.  
PR 35-AUG-1999; 99US-0147260.  
PR 36-AUG-1999; 99US-0147303.  
PR 36-AUG-1999; 99US-0147416.  
PR 36-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 15-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.





PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136792.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145046.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145213.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147049.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147142.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148505.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151940.  
PR 07-SEP-1999; 99US-0152243.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

```
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match: 10.0%; Score 174.5; FR 21; Length 400;
Best Local Similarity 24.7%; Pred. No. 4.5e-05;
Matches 90; Conservative 41; Mismatches 113; Indels 121; Gaps 14;

QY 63 GAENSEODIPSQL--LKMPLNPONGPLMV-----NGAEKGVLAIVIESMLPRGVDPYV 115
DB 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 2 GGVIKNDASADIKNEVLTTHLSGPIRVVDEEGVAARAGDULAVEICNIGLPQDQWG 61
QY 116 I-----CA-----MIP--HFGGLT--GTDLTAMLNDP 138
DB 62 FTGSFDRENGGFLTDHPCATKAIWYEGYAYSPQIPGVRRPGLTHGVIGIA---P 117
QY 139 LPKVRMIKLSKRVYSKSH-----TLPIKPHIGTUSVSP-----IDSINSI 182
DB 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 118 SNELMRT-----WNDEROLEESGVESLTCEVWHQRP:ACLPITKGCILGNIIEG 168
QY 183 TPO-----NHGGNDVPDIPGSGITIVLPVAPGGRIFGDHACOGUSET 227
DB 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 169 TPWERTANEAARTIPRENGCNDKNSRGSKIYLPVFEVAGANLSTGJMHFSQGDGFI 228
QY 228 --CGTAVEFASITIKVDLKNQLSW-----PRMENAENIMSIGSARP----- 269
DB 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 229 SFCC-AIEMSGFLELKEIIRNCMQEYLPMPGPTPLRVNPFIFEIGVPRFSENLYV:FGI 287
QY 270 -----LEDAIRAYRDLIYWLVE---DFGEQNDAYMLISOCG-KVRLGNWVCPKY 316
DB 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 288 SVDESGKQHYLDAT-VAYKRAVLNAIDYLFKFGYSKEQVYLLSCCPGRLSGIVDSPN 346
QY 317 TVGAM 321
DB 347 AVATL 351

RESULT 12
AAW40275
ID AAW40275 standard; Protein; 43 AA.
AC AAW40275;
XX
XX 16-JUN-1998 (first entry)
DT
DE K. oxytoca R-specific amidohydrolase peptide 113-3.
KW R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
KW 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
XX
XX Klebsiella oxytoca.
CS
XX
XX W05801568-A2.
PN
XX
XX 15-JAN-1998.
PD
XX
XX 10-JUL-1997; 97WO-EP03670.
ZF
XX
XX 03-MAR-1997; 97CH-0000500.
PR
XX 10-JUL-1996; 96CH-0001723.
XX
XX (LONZ ) LONZA AG.
PA
XX
XX Brieden W, Naughton A, Robins K, Shaw N, Tinschert A;
PI Zimmermann T;
FI
XX
XX WPI: 1998-101063/09.
DR
XX
XX N-PSDB: AAV10452.
XX
XX Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation;
PT - by stereoselective hydrolysis of corresponding racemic amide using
PT microorganism or derived enzyme, used as drug intermediate
XX
XX Example 10.2; Page 29; 68pp; German.
PS
```

```
XX
XX AAW40264-W40275 are peptide fragments of an R-specific amidohydrolase
XX isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
XX microorganism to utilize 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
XX as its sole nitrogen source. This amidohydrolase is used in a process
XX for preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl
XX propionamide which is cheaper than prior art optical resolution of the
XX racemate using dimethoxy strychnine or (S)-(-)-alpha
XX -methylbenzylamine.
XX
XX Sequence 33 AA;

Query Match: 8.5%; Score 148; DB 19; Length 33;
Best Local Similarity 87.9%; Pred. No. 5.9e-08;
Matches 29; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 110 GVDPYGICAMIPHEGGLTGDTLTAMLNDPLPEK 142
DB 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 1 GVDPYGIEAMIPHEGGLTGDTLTAMLNDQLQPK 33

RESULT 13
ABB93277
ID ABB93277 standard; Protein; 307 AA.
AC ABB93277;
XX
XX 31-MAY-2002 (first entry)
DT
DE Herbicidally active polypeptide SEQ ID NO 2488.
KW Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
XX
XX W0200210210-A2.
PN
XX 07-FEB-2002.
PD
XX 26-AUG-2001; 2001WO-EP09892.
XX
XX 28-AUG-2001; 2001WO-EP09892.
PR
XX (FARB ) BAYER AG.
PA
XX Tie-Jen K. Weidner M;
XX
XX WPI: 2002-269010/31.
DR
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 2488; 261pp - Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 307 AA;

Query Match: 7.4%; Score 130; DB 23; Length 307;
Best Local Similarity 23.9%; Pred. No. 0.00012;
Matches 62; Conservative 26; Mismatches 95; Indels 76; Gaps 11;

QY 26 TEEMQKEFHYTIGYSTPVLITIEPDRIIVDTDAFEGL---NSEQDIPSQLKMPFLN 82
```

```

DB          1 EPHYTIGPSTPVLTIERGDR 21
|||||:|||||||:|||||:|||||
RESULT 15
AB849-56
ID   AB849:50 standard; Protein; 238 AA.
XX AC      AH549156;
XX ET      05-FEB-2002 (first entry)
XX DE      Listeria monocytogenes protein #1862.
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX CC      Listeria Monocytogenes.
XX WN       WQZ00177335-A2.
XX PD       18-OCT-2001.
XX LP       13-APR-2001; 2001WO-FR01118. *
XX RP       17-APR-2003; 2003FR-0064629.
XX TA       (INSP ) INST PASTEUR.
XX PF       Buchrieser C, Frangelet E, Couve E, Kushiok C, Psihi H, Dehoux P,
PF         Dussurget O, Chetouiati F, Nedjari H, Glaser P, Kunst F, Gossart P,
PI         Daniels J, Goebel W, Kneft J, Kuhn M, Ng E, Vazquez-Boland JA,
PI         Dominguez-Bernal G, Garrido-Garcia R, Tierrez-Martinez A, Amend A,
PI         Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,
PI         Perez-Staez J, Buquero F, Garcia del Portillo F, Gomez-Lopez N,
PI         Macuenio E, De Pablo B, Weiland J, Kaerst U, Entian K, Hauf J,
XX         Rose M, Voss H;
XX WP       2002-010914/01.
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PI related polypeptides -
XX CA       Claim 6; SEQ ID NO 1863; 192pp; French.
CC The present invention relates to the genome sequence of Listeria
CC monocytogenes BGDc (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC expressed from the genome sequence of the present invention. Proteins
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ       Seq.ence    238 AA;

Query Match           6.1%; Score 107; DB 23; Length 238;
Best Local Similarity 22.2%; Pred.No. 0.019;
Matches 55; Conservative 24; Mismatches 75; Indels 94; Gaps 9;

```

Db 51 HKLKSPKRLILTHAGDHIGSLIAVKMAAPEAL- ----- VMIG 48  
QY 92 GAEKGDVLAVYIESMLPROVDPYGICAMIPHEGSGTDLZAMLADHILPEKVRMIKLPSE 151  
Db 89 SREK-----LAVETKEIYAFEAQK-----LKGYSDDLFLVRIDQI:KMGD 129  
QY 152 KYVWSKRHTLPYKPHIGTILSVSPE:DSINSLTPDNHGGNMVPIGPGSITYFLVRAFGG 211  
Db 130 MV-----GSLI-----IDTPGHIPGISTFFDEK--NG 155  
QY 212 RLFIGDAHACQGDGETCGTAVAFASITLIKVDLIKKNWOLSWPRMENAENIKSIGSARULE 271  
Db 156 HLFVGDLFOTRGGAACGE-----KRWLFFPFAMGSDPLPTSLASAENLQ 200  
QY 272 --DARIA 277  
Db 201 LFDVTEIA 208

Search completed: July 3, 2003, 15:47:16  
Job time : 84 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 15:45:52 : Search time 24 seconds  
(without alignments)  
37,181 Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1748

Sequence: 1 MKWLPSIMAKRQVACGRKP.....GNVDPKITYGAMCNKNDLV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942222 residues

Total number of hits satisfying chosen parameters: 26257

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	177.5	10.2	575	US-08-922-865-2	Sequence 2, Appli
2	177.5	10.2	575	US-09-510-949-2	Sequence 2, Appli
3	91.5	5.2	1848	US-08-296-791-6	Sequence 6, Appli
4	91.5	5.2	1849	PCT-US95-10661A-6	Sequence 6, Appli
5	82.5	4.7	500	US-08-926-8429-64	Sequence 64, Appli
6	82.5	4.7	857	US-08-778-113-2	Sequence 2, Appli
7	82.5	4.7	958	US-08-583-5625-2	Sequence 2, Appli
8	82.5	4.7	1752	US-09-556-877-180	Sequence 180, App
9	82.5	4.7	1752	US-09-620-412C-180	Sequence 30, App
10	82	4.7	540	US-09-177-165A-30	Sequence 42, App
11	81.5	4.7	1170	US-08-476-062A-42	Sequence 42, App
12	81.5	4.7	1170	PCT-US96-0131A-42	Sequence 42, App
13	81	4.6	1342	US-07-976-895-4	Sequence 4, Appli
14	81	4.6	1342	US-08-473-119-4	Sequence 4, Appli
15	81	4.6	1342	US-08-475-352-4	Sequence 4, Appli
16	81	4.6	1343	US-08-475-352-4	Patent No. 513864
17	80	4.6	628	US-07-952-853-22	Sequence 22, Appli
18	80	4.6	528	US-08-914-848-22	Sequence 22, Appli
19	79.5	4.5	1544	US-09-413-814-45	Sequence 45, Appli
20	79	4.5	374	US-09-306-446C-2	Sequence 2, Appli
21	79	4.5	1342	US-08-484-438-9	Sequence 2, Appli
22	79	4.5	1711	US-08-368-822C-10	Sequence 10, Appli
23	79	4.5	1711	US-08-582-776C-10	Sequence 10, Appli
24	79	4.5	1711	US-08-434-8315-10	Sequence 10, Appli
25	78.5	4.5	375	US-09-171-337A-7	Sequence 7, Appli
26	78	4.5	383	US-08-391-916A-4	Sequence 4, Appli
27	78	4.5	526	US-08-391-916A-4	Patent No. 5258290

26	77.5	4.4	1170	2	US-08-789-078-2	Sequence 2, Appli
29	77.5	4.4	1170	2	US-08-752-633-2	Sequence 2, Appli
30	77.5	4.4	1170	5	PCT-US95-04886-2	Sequence 2, Appli
31	77	4.4	565	4	US-09-142-623-11	Sequence 11, Appli
32	77	4.4	575	1	US-08-403-866-7	Sequence 7, Appli
33	76.5	4.4	215	1	US-08-107-684B-10	Sequence 10, Appli
34	76.5	4.4	215	1	US-08-107-684B-11	Sequence 11, Appli
35	76	4.3	456	4	US-09-268-354-21	Sequence 21, Appli
36	76	4.3	629	4	US-09-134-001C-4354	Sequence 4354, Ap
37	76	4.3	1065	3	US-08-630-172-9	Sequence 9, Appli
38	76	4.3	1065	4	US-09-375-419-9	Sequence 9, Appli
39	76	4.3	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
40	76	4.3	1222	2	US-08-682-517-15	Sequence 15, Appli
41	76	4.3	1252	2	US-08-682-517-9	Sequence 9, Appli
42	76	4.3	1346	3	US-09-320-878-4	Sequence 4, Appli
43	76	4.3	1346	4	US-09-105-537-37	Sequence 37, Appli
44	76	4.3	11877	4	US-09-105-537-6	Sequence 6, Appli
45	75.5	4.3	354	4	US-09-500-569-12	Sequence 12, Appli

## ALIGNMENTS

RESULT 1  
US-08-922-865-2  
Sequence 2, Application US/08922865  
Patent No. 6090616  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: MICROBICAN™, LACTAMASE ENZYME OBTAINED  
TITLE OF INVENTION: THEREFROM, AND THEIR USE  
NUMBER OF SEQUENCES: 2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Pateulin Release #1.0, Version #1.39 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/922,865  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-922-865-2

Query Match	10.2%	Score 177.5	DB 3	Length 575
Best Local Similarity	24.0%	Pred. No. 3e-11		
Matches	99	Conservative	33	Mismatches 85; Indels 101; Gaps 12
QY	39	PYSTP-----	VLTIPEGRIIVDTRDAEGAI---NSEQDIPS-QLKMPFLN 82	
DB	44	PDYQPVHNRKHPDI PMAVVVEPGAEFKLETVDWTGGALKNDSDADVDLSTVHFL- 102		
QY	83	PQNSPFWNCAEKSDYLAV-----	YIESMLPRGV-----D 112	
DB	103	--SGPVGVKGQSPDLVLLVDLDIGARDSDINMGNGFFSKQNGGFLDEHFLAOKSIWD 160		
QY	113	PVSGICAMIPH-----FCG:GTGDTATMLNDPIPEKVRMTKLDSEKYWSKRHTLPYKPHI 167		
DB	161	FHGMFTKSRHPIGVNFAGLIHFLIGLCFLDP-----KML-----ASNRET----- 202		
QY	168	GLTSLVSPF-----	DSINSLTPDNHGGNMDVDPDIPGS 199	
DB	203	GLIATDPDRPIGLANPNATIAHMGOMOGGEARDKAAAGARTVPPRHGNGCNDKLSRG 262		
QY	200	SITYPYLVANGKRLFGIDAHACGGGEGIC-GTAVEFASITITIKVDLIK 246		
DB	263	SRVFFVYVVGAGLSVCDLHFSOGDGEITFWGPIEMPGVHMKVSLIK 310		

## RESULT 2

US-09-510-949-2  
: Sequence 2, Application US/09510949  
: Patent No. 6423522  
: GENERAL INFORMATION:  
: APPLICANT: St. Geme III, Joseph W.  
: TITLE OF INVENTION: MICROORGANISM, LACTAMASE ENZYME OBTAINED  
: FROM THE INVENTION: THEREFROM, AND THEIR USE  
: NUMBER OF SEQUENCES: 2  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.3C (LPO)  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/510.949  
: FILING DATE:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 08/922,865  
: FILING DATE:  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 575 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
US-09-510-949-2

Query Match 10.2%; Score 177.5; DB 4; Length 575;  
Best Local Similarity 24.0%; Pred. No. 3e-11;  
Matches 69; Conservative 33; Mismatches 85; Indels 101; Gaps 12;  
QY 39 PYSTP-----VLTIEGDRIVITDRAFEAI---NSEQIPIS-QLLKMFELN 82  
DB 44 PYDNQVNRWHPDIPMAVWEPGAEFKLEYDTGTGAIKNDQSAEDVRVDLSTVIFL- 102  
QY 83 PONGIMVNGAEKGDVLAV-----YTESKMPKSV-----D 112  
DB 103 --SGPVGKQAGQGLLVYDLDIGARDSDLWNGCFKSKONGGFIIDEHFPLACKSMD 160  
QY 113 PYGICAMIPH-----FGGLGTDLTAMLDPLPEKVMIKLDSEKVVYVSKRELIPYKPH: 167  
DB 161 FPGMTKSHHIVGVNFAGLIHPLGLGLPDP-----KYL-----ASWNERET----- 212  
QY 168 GTLSVSPF-----TDSINSLTPDHGGMDVPTICPG 199  
DB 203 GLIATDPDRIPGIANPPNATTAAHMQSGEARGKAAAEAGATVPPRSHGGNCDTKLSRG 262  
QY 209 SLTYPLVRAPGGRPLFGDAHACGGGEGIC-GTAVEFASITIKVQDLAK 246  
DB 263 SRVFFPVYVVGAGLSVGLDHFSGGCGELTFWGPTEMPGVVWVKVSLIK 310

## RESULT 3

US-08-296-791-6  
: Sequence 6, Application US/08296791  
: Patent No. 6245337  
: GENERAL INFORMATION:  
: APPLICANT: St. Geme III, Joseph W.  
: TITLE OF INVENTION: Haemophilus Adherence and Penetration  
: FROM THE INVENTION: Haemophilus Adherence and Penetration  
: NUMBER OF SEQUENCES: 9  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
: STREET: 4 Embarcadero Center, Suite 3400  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: United States  
: ZIP: 94111-4187  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: SOFTWARE: PatentIn Release #1.0, Version #1.25

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/296.791  
FILING DATE: 25-AUG-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Treccartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-59941/RF\*/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1848 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
US-08-296-791-6

Query Match 5.2%; Score 91.5; DB 4; Length 1848;  
Best Local Similarity 19.1%; Pred. No. 1.4;  
Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 20;  
QY 35 YTGIPYTPVLIIEPGD---RIIVDTRDAFE-GAINSE-QDIPSOLLKMPFLNPONGPI 88  
DB 17 YALPYTEAALVRDDVDYQIFRDEFAENKGFSGVATNVEVRKKNQSL-----GSA 67  
QY 89 MVNGAEKGDVJAVYIESMLPRGVDPY-----GICAMIPHFGLGTGDTLAMLNDP 138  
DB 68 LPNGIPMIDFSDVDKRIATLVNPQYVGVGVKHSVNGSEL--HFGNLNGN-----MNG 120  
QY 139 LPKVRMIKLDSEKVVYVSKRELIP-----YKPHIGTILSVSPRID 177  
DB 121 NAKSHROVSSSENYITVYKNNFTENVTSTKEEQDAQRREDYTPRLDKFVT--EVA 178  
QY 178 SINSITPDHGGMDVDPDIPGSGIYTP-LVRAPGGRLEIGDAHACGGGEGICGTAVEFAS 236  
DB 179 PIEASTANNKGEYNSD-----KYPAFVRLSGTGFLYKKS-----KYQL 220  
QY 217 ITTK---VDLKKWQLSWPME-----NAENIMSISGARPLE---DA 273  
DB 221 ILTEKOKGNIERNMDVGGDNLDLVGNAYTYGIAGTPYKVNHNENGLIGFGNSKEHSDP 280  
QY 274 TRIAYRD-----LIY-----EDF-----GFQWDAYMLLSQ 302  
DB 281 KGLISODPLINAVLGDGSGPLFVYDREKCKWFLGSDYDFWAGYNKKSQWENYI----- 335  
QY 303 CGKVPGLGNWDPKTYVGAMLNKN 325  
DB 346 --KHEFAEKYQOYSAGSLHGSN 356

## RESULT 4

US-0951-10261A-6  
: Sequence 6, Application PC/T095951C661A  
: GENERAL INFORMATION:  
: APPLICANT: Washington University, et al.  
: TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
: FROM THE INVENTION: Haemophilus Adherence and Penetration Protein  
: NUMBER OF SEQUENCES: 9  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
: STREET: 4 Embarcadero Center, Suite 3400  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: United States  
: ZIP: 94111-4187  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```
/
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/10661A
/ FILING DATE: 16-AUG-1995
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 68/296,791
/ FILING DATE: 25-AUG-1994
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Itecartin, Richard F.
/ REGISTRATION NUMBER: 31,601
/ REFERENCE/DOCKET NUMBER: PP-59941/RFT
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 781-1989
/ TELEFAX: (415) 398-3249
/ TELEFAX: 910 277299
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1848 amino acids
/ TYPE: amino acid
/ TOPOLOGY: unknown
/ PCT-US95-10661A-6

Query Match 5.2%; Score 91.5; DB 5; Length 1848;
Best Local Similarity 29.1%; Pred. No. 1.4;
Matches 73; Conservative 48; Mismatches 127; Indels 135; Gaps 27;

QY 35 YHIGYSTPLVTIEPGD-----KIIVDTDAFE-GAINSE-QDIPSQLKMPFLNTQNEP; 83
DB 17 YALTPYTEAALVRDDVDYQ;FSDFAENKGFESVGAINVEVRKKNQS;-----GSA 67
QY 89 MYNGAEKGDVJAVYTESMLPRGVDPY-----GICAM;PIPGGLTGDTLAMLNDP 138
DB 68 LPLNGIPMDIFSVDVDEKRIATILNPQYVGVKHSNGVSEL--HFGNLGN-----MNG 120
QY 139 LPEKVRMIKIDSEKVMKRRHTLP-----YKPHIGTLSVSRPID 177
DB 121 NAKSHRDVSSSEENRYTYVEKKNFENTSTKEFQDAQRREDYMPRLXEV--EVA 178
QY 178 SINSUTPDNNGNMDVPDIPGSIYTP-LVRAPGRLRFIGAHACQGGGECGTAVEFAS 236
DB 179 PIEASTANNKGEYNNSD-----RYPAFVRLGSGTQFYKKG;-----RVQL 220
QY 237 ITTIK---VDLIKNSLWPRME-----NAENIMTSGSARPLE---DA 273
DB 221 ILTEKDKGNLLRNWDVGDGDLVGNAYTYGIAGTPYKVNHNHNLIGFNSKSEISDP 280
QY 274 TRIAYRD-----LIY-----WLV---EDF-----GPFQWDAYMLLSQ 302
DB 281 KGIISQDPLTNYAVIGDSGLFVYDREKGGKFLGSDYDFWAGYNKKSQEWNIY----- 335
QY 303 CGKVRLGNNMVDPKYTVGAMLNKN 325
DB 336 --KHEFAEKIYQOYSAGSGLIGSN 356

RESULT 5
US-08-926-842B-64
Sequence 64, Application US/08926842B
Patent No. 6030807
GENERAL INFORMATION:
APPLICANT: Sa-No. 6030807uira, Isabel
APPLICANT: de Lencastre, Herminia
TITLE OF INVENTION: HIGHLY REGULABLE PROMOTER FOR HETEROLOGOUS GENE
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

/
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US 08/926,842B
/ FILING DATE: 10-SEP-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Jackson Esq., David A.
/ REGISTRATION NUMBER: 26,742
/ REFERENCE/DOCKET NUMBER: 603-1-089 N
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201 487-5800
/ TELEFAX: 201 343-1684
/ TELEFAX: 133521
/ INFORMATION FOR SEQ ID NO: 64:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 500 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULAR TYPE: protein
/ ORIGINAL SOURCE:
/ ORGANISM: Bacillus subtilis
/ FEATURE:
/ OTHER INFORMATION: /product= "abfa"
/ US-08-926-842B-64

Query Match 4.7%; Score 82.5; DB 3; Length 500;
Best Local Similarity 25.4%; Pred. No. 1.7;
Matches 31; Conservative 32; Mismatches 75; Indels 43; Gaps 13;

QY 142 EKVMTIKLUSE-----KVYMSKRHTLPYKPHIGTLSVSPFIDSINSIPDNHGGNMD 192
DB 2 KRAMI-VDKEYKIGVDKRIYGS-----FIEHMR-AVVEGIVPEORPEADEGFRKD 53
QY 193 VPDGPGSITYLVRAPGRLRFIGAHACQGGGECG--TAVEFASITIKVOLKNNQL 250
DB 54 VQSL-IKELOVPIIRYPGGN-FLSGYNWEDGVGPVNRPRRLDLAWQTETNEVGTNEFL 111
QY 251 SWPRMENAENINSIG-SARPLEDATRIAYRDLI-----YW--LVEDFGFEQWDAYM 298
DB 112 SWAKKVNTEVNNAVNIGIRGID-----AARNLVEYCNHPKGSYWSDLRRSHGYEQ--PYG 164
QY 299 LLSQCKVRLGNMVDPKYTVG 319
DB 165 IRTWC-----IGNEMGPMQIG 181

RESULT 6
US-08-779-113-2
Sequence 2, Application US/08779113
Patent No. 5948891
GENERAL INFORMATION:
APPLICANT: Staunton, Donald E.
APPLICANT: Harris, Edith S.
TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
TITLE OF INVENTION: Binding
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gorstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```



```

: APPLICATION NUMBER: US/08/779,113
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Greta E. No. 594889and
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 27866/33773
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-474-6300
: TELEFAX: 312-474-0448
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 857 amino acids
: TYPE: amino acid
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
US-08-779-113-2

Query Match 4.7%; Score 82.5; DB 2; Length 857;
Best Local Similarity 20.3%; Pred. No. 4.3;
Matches 48; Conservative 31; Mismatches 85; Indels 73; Gaps 9;

QY 5 EESIMAKR-----GVGAGRKPVTHLTETEMOKEHYITGIPYSTPVLTIPEGDR----- 52
Db 204 DEPLLKRLPNHHIGISFIPREVGHEHLSVSKNKNHVNANSPVIMVQSEIGDARRAKVY 263
QY 53 -----IIVDTRDAFEGAINSEODIPSQLKMPFLNPONGP-MWNGAEKGDV 98
Db 264 GRGLSEGRTEFMSDFIVDTRDAGYGGIS-----LAVEGSPSKVDI 302
QY 99 LAVVIESMLPRGVDPYIGICAMIPHFGLGTGDL--TAMLNDDPLPEKVRMIKLDSEKVVYS 156
Db 303 QTEDELED-----GTC-KVSYFPTVPGVYIVSTKFADEHVPGPSFTVVKISGEG---- 348
QY 157 KRHTLPYKPHIGTLSVSPEDSINSLTDPNHHGNMDVDPDIPGSGITYPLVRAPGGRL 213
Db 349 -----RVKESIRTSRAPSVATVGSIC----DLNLKIPETNSSDMS-AHVTSPSGRV 395

RESULT 7
US-08-583-562B-2
: Sequence 2, Application US/08583562B
: Patent No. 5922570
: GENERAL INFORMATION:
: APPLICANT: Staunton, Donald
: APPLICANT: Harris, Edith
: TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
: TITLE OF INVENTION: Binding
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 233 South Wacker Drive, 6300 Sears Tower
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/583,562B
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Williams Jr., Joseph A.
: REGISTRATION NUMBER: 38,659
: REFERENCE/DOCKET NUMBER: 27866/33033
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-474-6300
: TELEFAX: 312-474-0448
: INFORMATION FOR SEQ ID NO: 2:

```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 858 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-583-562B-2

Query Match 4.7%; Score 82.5; DB 2; Length 858;
Best Local Similarity 20.3%; Pred. No. 4.3;
Matches 48; Conservative 31; Mismatches 85; Indels 73; Gaps 9;

QY 5 EESIMAKR-----GVGAGRKPVTHLTETEMOKEHYITGIPYSTPVLTIPEGDR----- 52
Db 204 DEPLLKRLPNHHIGISFIPREVGHEHLSVSKNKNHVNANSPVIMVQSEIGDARRAKVY 263
QY 53 -----IIVDTRDAFEGAINSEODIPSQLKMPFLNPONGPIMVNGAEKGDV 98
Db 264 GRGLSEGRTEFMSDFIVDTRDAGYGGIS-----LAVEGSPSKVDI 302
QY 99 LAVVIESMLPRGVDPYIGICAMIPHFGLGTGDL--TAMLNDDPLPEKVRMIKLDSEKVVYS 156
Db 303 QTEDELED-----GTC-KVSYFPTVPGVYIVSTKFADEHVPGPSFTVVKISGEG---- 348
QY 157 KRHTLPYKPHIGTLSVSPEDSINSLTDPNHHGNMDVDPDIPGSGITYPLVRAPGGRL 213
Db 349 -----RVKESIRTSRAPSVATVGSIC----DLNLKIPETNSSDMS-AHVTSPSGRV 395

RESULT 8
US-09-556-877-180
: Sequence 180, Application US/09556877
: Patent No. 6432916
: GENERAL INFORMATION:
: APPLICANT: Probst, Peter
: APPLICANT: Bhatia, Ajay
: APPLICANT: Skeiky, Yasar
: APPLICANT: Fling, Steve
: APPLICANT: Maisonneuve, Jeff
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
: TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
: FILE REFERENCE: 210121-469C5
: CURRENT APPLICATION NUMBER: US/09/556,877
: CURRENT FILING DATE: 2000-04-19
: NUMBER OF SEQ ID NOS: 305
: SOFTWARE: FastSeq for Windows Version 3.0/4.0
: SEQ ID NO 180
: LENGTH: 1752
: TYPE: PRI
: ORGANISM: Chlamydia
US-09-556-877-180

Query Match 4.7%; Score 82.5; DB 4; Length 1752;
Best Local Similarity 23.1%; Pred. No. 14;
Matches 49; Conservative 26; Mismatches 60; Indels 77; Gaps 10;

QY 40 YSTPVLTIPEGDRRIIVDTRDAFE--GAINSEODIPSQLKMPFLNPONGPIMVNGAE--- 94
Db 271 YAKGDLTIADSOQVLP-SINKATKDGGAIFAERDV-----SFENITSLKVOINGAEKKG 323
QY 95 -----KGDVL-----AVIESMLPRGVDPYIGICAMIPHFGLGTGD 130
Db 324 GAIYAKGDLSTIOSSKUSFNSNYSKGGGALYVE-----GGINFQD 364
QY 133 LTAMLNDDPLPEKVRMIKLDSEKVRHTLPYKPHIGTISVSPEDSINSLTDPNHHGN 190
Db 365 L-----EEIR-IKYNKAGTTEIKKIL---PSLKAQASAGNADAWASSPQSGGA 411
QY 191 MDVPDIPGPGSI-----TYPLVRAPGGRL 214
Db 412 TTVDSDSGDSSGSDSDTSETVP-VTAKGGGLY 442

RESULT 9

```

```
US-09-620-412C-180
: Sequence 180, Application US/09620412C
: Patent No. 6448234
: GENERAL INFORMATION:
: APPLICANT: Steven P. Fling
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
: FILE REFERENCE: 210121.469C7
: CURRENT APPLICATION NUMBER: US/09/620.412C
: CURRENT FILING DATE: 2000-07-20
: NUMBER OF SEQ ID NOS: 363
: SOFTWARE: FASTSEQ for Windows version: 3.0/4.0
: SEQ ID NO 180
: LENGTH: 1752
: TYPE: PRT
: ORGANISM: Chlamydia
US-09-620-412C-180

Query Match      4.7%  Score 82.5; DB 4; Length 1752;
Best Local Similarity 23.1%; Pred. No. 14;
Matches 49; Conservative 26; Mismatches 60; Indels 77; Gaps 10;

QY 40 YSTPLTIEPDRIVITRCAFE--GAINSEODIPSOILKMPFLNPONGPIWNGAP--- 94
DB 271 YAKGLTIADQEVLFSTINKAKKGGALFAEKW-----SFNTISLAVQNGAEKQ 123
QY 95 -----KGNVL-----AVYIESMLPRGVWYVG:CAMIFHFGGLTGTD 130
DB 324 GATYAKGRLSTOSSKQSLFNSYNSKQGGALYVF-----GQINEQ 364
QY 131 LTAMLNDFLPKVMKILDSKVYWSKRHTLPYKPHIGTISVGEFDSINSLTFHNGGN 140
DB 305 L-----SEIR-IKYKAGTETTKITL---PS:KAAAGACNADAKWASSSTGSGSQA 411
QY 191 MDVPIGRPSI-----TYPLVRAPGGR:F 214
DB 412 TVSDSGDSGSDSDTSETVP-VTAKGGSIY 442

RESULT 10
US-09-177-165A-30
: Sequence 30, Application US/09177165A
: Patent No. 6426205
: GENERAL INFORMATION:
: APPLICANT: Tyers, Mike
: APPLICANT: Willemis, Andrew
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
: FILE REFERENCE: 11757.1025U1
: CURRENT APPLICATION NUMBER: US/09/177.165A
: CURRENT FILING DATE: 1998-10-22
: PRIOR APPLICATION NUMBER: 60/092,443
: PRIOR FILING DATE: 1998-07-10
: PRIOR APPLICATION NUMBER: 60/063,254
: PRIOR FILING DATE: 1997-10-24
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 30
: LENGTH: 640
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-30

Query Match      4.7%  Score 82; DB 4; Length 640;
Best Local Similarity 21.1%; Pred. No. 3;
Matches 39; Conservative 23; Mismatches 65; Indels 56; Gaps 7;

QY 62 NPONGPIWNGAEKGDVLAVYIESMLPRGVDPYGLICAM:PEFSULTGTGIDLFAM:AEPLPE 141
DB 490 SPQDPTMTDGADESDTSPNEQ:VLDK-----NLPVPT 523
QY 142 KYRMKILDSKVYWSKRHTLPYKPHIGTISVSPEDINSITPKN-----HCGNMVVP 194
```

```
DB 524 HLLSCGLDNTIKLWZVKTGKCIQTQF-----HVEGVWDIAADNFRIISGSHDGIKW 577
QY 195 DIGPISITVPLVRAPGGR:L-----FIGD-----AHACQGDGEICGTAVEFASIT 239
DB 578 DLQSGKCMHIF---NGRRLQRETOHTOTQSLGDKVAPIACVCIGDSE-CFSGDEFCVKM 633
QY 240 IKVDL 244
DB 634 YKFDL 638

RESULT 12
US-08-476-062A-42
: Sequence 42, Application US/08476062A
: Patent No. 5877275
: GENERAL INFORMATION:
: APPLICANT: Ardaout, M. Amin
: TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
: TITLE OF INVENTION: RESPONSES WITH BETA2 INTEGRINS
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/476.062A
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/216.081
: FILING DATE: 21-MAR-1994
: APPLICATION NUMBER: 07/637.830
: FILING DATE: 04-JAN-1991
: APPLICATION NUMBER: 07/539.842
: FILING DATE: 18-JUN-1990
: APPLICATION NUMBER: 07/212.573
: FILING DATE: 28-JUN-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Freeman, John W.
: REGISTRATION NUMBER: 29,066
: REFERENCE/DOCKET NUMBER: 00786/068003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 42:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1170 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: US-08-476-062A-42

Query Match      4.7%  Score 81.5; DB 2; Length 1170;
Best Local Similarity 22.7%; Pred. No. 9.4;
Matches 70; Conservative 35; Mismatches 109; Indels 95; Gaps 18;

QY 14 VAGAKKPYTHHLEKQKEFH-YIIGPYSTVPLTIEPDRIVITRCAFECAINSEQDIP 72
DB 284 IGIGK-----HFGTKESQETLHKFAKPKASE-----FVKILOIFE----- 318
QY 73 SOLLKMPFLNPONGPIWNGAEKGDVLAVY:E-----SMLPRGVDPYGLICAM:PHFG 125
DB 319 --KLDFLPIEROKKIYVIEGTSSKQDLTSFNMELSSGSLASLSRGHAVVGAKWAGG 376
```

QY 126 LTGDTLAML-----NDLPEKVRMIKLDSEKVVW--SKRHT- ----LPEKRIHT 170  
DB 377 F--LDLKAQDQ:TFIGNEPHTEVRAGY:G-YTVIWLNSRCKTS:LASASAPYQHMGRV 433  
QY 171 SVSPE-----IDS:NLSDPNH:GGMNDVPDIG:PGSITYPIVRAP-----GGR 212  
DB 434 LLFOEGGQGHWSQVQTHGTQ:GSGYFGGELCGVDVQDQGETELLIGAPLFYGEQGRG 493  
QY 213 LFI-----GUAHACQGD:GEICGTAVEFAISITIKVDLIKNNWLSWPKENAINMS 263  
DB 494 VFIYORQLGFEVSELOQDPGYPLG---RFGFAITALT:DI-----NGDGIYD 538  
QY 264 IGSARPLED 272  
DB 539 VAVGAPLEE 547

## RESULT 12

PCT-US96-01314-42

: Sequence 42, Application PC/70US9601314

: GENERAL INFORMATION:

: APPLICANT: M. Amin Arnaout

: TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN

: NUMBER OF SEQUENCES: 78

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Fish &amp; Richardson P.C.

: STREET: 225 Franklin Street

: CITY: Boston

: STATE: Massachusetts

: COUNTRY: U.S.A.

: ZIP: 02110-2804

: COMPUTER READABLE FORM:

: MEDIUM TYPE: 3.5" Diskette, 1.44 MB

: COMPUTER: IBM PS/2 Model 502 or 55SX

: OPERATING SYSTEM: MS-DOS (Version 5.0)

: SOFTWARE: WordPerfect (Version 5.1)

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: PCT/US96/01314

: FILING DATE: 30-JAN-96

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 08/380,167

: FILING DATE: 30-JAN-95

: ATTORNEY/AGENT INFORMATION:

: NAME: John W. Freeman

: REGISTRATION NUMBER: 29,045

: REFERENCE/DOCKET NUMBER: 95786/257001

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (617) 542-5070

: TELEFAX: (617) 542-5966

: TELEX: 200154

: INFORMATION FOR SEQ ID NO: 42:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1170

: TYPE: amino acid

: STRANDEDNESS:

: TOPOLOGY: linear

: PCT-US96-01314-42

## Query Match

Best Local Similarity 4.78; Score 81.5; DE 5; Length 1170;

Matches 70; Conservative 35; Mismatches 109; Indels 95; Gaps 18;

QY 14 VQAGRKPYTHLTDMQKEFH-YTIGPYSTPVLIIPEGDRRIIVDIRDAFEGAINSECDIP 72  
DB 284 IGIGK---HFOTRESQET:HKFASKPASE-----FVKILSTEE----- 318  
QY 73 SOLLKMPNLQNGPIWNGAEKGDVLAHYIE-----SMLPGVDYPIGICAMIFRFG 125  
DB 319 --KKDLDF:EKOKIYIVIEGTSKQDLTSFNMLSSG:ISAD:SRGAVGVAVAKQWAGG 376  
QY 126 LTGDTLAML-----NDLPEKVRMIKLDSEKVVW--SKRHT- ----LPEKRIHT 170

DB 377 F--LDLKAQDQ:TFIGNEPHTEVRAGY:G-YTVIWLNSRCKTS:LASASAPYQHMGRV 433  
QY 171 SVSPE-----IDS:NLSDPNH:GGMNDVPDIG:PGSITYPIVRAP-----GGR 212  
DB 434 LLFOEGGQGHWSQVQTHGTQ:GSGYFGGELCGVDVQDQGETELLIGAPLFYGEQGRG 493  
QY 213 LFI-----GUAHACQGD:GEICGTAVEFAISITIKVDLIKNNWLSWPKENAINMS 263  
DB 494 VFIYORQLGFEVSELOQDPGYPLG---RFGFAITALT:DI-----NGDGIYD 538  
QY 264 IGSARPLED 272  
DB 539 VAVGAPLEE 547

## RESULT 13

US-07-978-895-4

: Sequence 4, Application US/07978895

: Patent No. 5480968

: GENERAL INFORMATION:

: APPLICANT: Kraus, Matthias H.

: APPLICANT: Aaronson, Stuart A.

: TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE

: TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND

: TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THERETO

: NUMBER OF SEQUENCES: 12

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Suite 400

: STREET: 133 Carnegie Way, N.W.

: CITY: Atlanta

: STATE: Georgia

: COUNTRY: U.S.A.

: ZIP: 30303

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatib.e

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/07/978,895

: FILING DATE: 19921110

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 07/444,436

: FILING DATE: 01-DEC-1989

: ATTORNEY/AGENT INFORMATION:

: NAME: Percymar, David G.

: REGISTRATION NUMBER: 33,438

: REFERENCE/DOCKET NUMBER: 1414-028

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (404) 688-0776

: TELEFAX: (404) 688-9880

: INFORMATION FOR SEQ ID NO: 4:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1342 amino acids

: TYPE: AMINO ACID

: TOPOLOGY: linear

: MOLECULE TYPE: protein

US-07-978-895-4

## Query Match

Best Local Similarity 4.69; Score 81; DB 1; Length 1342;

Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY 8 IMAKRGVAGAKPYTHLTDMQKEFHVITGYPSTPVLIIPEGDRRIIVDIRDAFEGAINS 67  
DB 378 IKRSGPGIARCPHPGLINKKLEEV--LEPELDLDLDEAEEDNLATI-----TLGS 1029  
QY 66 EDD:PSQLKKMP-----FLNPQNGPIWNGAEKGDVLAHYIESML-----PRGVD--- 112  
DB 1040 ALSLPVGTNLNPRGQSILSPSGYMPMNOGNLGSQ---ESAVSGSSRCPRPVSLHP 1086  
QY 113 -PYGICAMIFHGGLGTDLTAMLDPLPEKVRMIKLDSEKVVW--SKRHT- ----LPEKRIHT 161

Db 1087 MPRGCLASSESGHVTGSEA-----ELQKVMCMCRSRSRSPRGRGDSAYHSORHSL 1140  
QY 162 -PYKPHICTLSVSP---EIDSINS-LTPDNH 187  
Db 1141 TPVTP-----LSPGGLSEEDVNGVMPDTH 1165

RESULT 14  
US-08-473-119-4  
: Sequence 4, Application US/08473119  
: Patent No. 5820859  
: GENERAL INFORMATION:  
: APPLICANT: Kraus, Matthias H.  
: APPLICANT: Aaronson, Stuart A.  
: TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE  
: TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND  
: TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREO  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Suite 400  
: STREET: 133 Carnegie Way, N.W.  
: CITY: Atlanta  
: STATE: Georgia  
: COUNTRY: U.S.A.  
: ZIP: 30303  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/473,119  
: FILING DATE: 07-JAN-1995  
: CLASSIFICATION: 424  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/978,895  
: FILING DATE: 10-NOV-1992  
: APPLICATION NUMBER: US 07/444,405  
: FILING DATE: 01-DEC-1989  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Perryman, David G.  
: REGISTRATION NUMBER: 33,438  
: REFERENCE/DOCKET NUMBER: 1414-028  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (404) 688-9880  
: TELEFAX: (404) 688-9880  
: INFORMATION FOR SEQ ID NO: 4:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1342 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-473-119-4

Query Match 4.6%; Score 81; DB 2; Length 1342;  
Best Local Similarity 24.6%; Pred. No. 13;  
Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY 8 IMAKRGVAGRKPVTHLLEENKQEFHYTIGPSTPV-TIEPGDRIIVDTKAEFGAINS 67  
Db 978 IKRESGPIAGCPGPHGLTNKKLEEVE--LEPELDLDLDEAEDNLATT-----TLGS 1029

QY 68 EQDIPSOLLKMP-----FLNPONGPTMVNGAEKGVLAIVIESML-----PRGVD--- 112  
Db 1030 ALSLPVGTNNRPGSSLLSPSGYMPMNGNLGSCQ---ESAVSGSSERCPRPVSLHP 1086

QY 113 -PYGICAMIPHFGLTGTLTAMLDPLPEKVRMKLDS-----EKVYWSKRHTL- 161  
Db 1087 MPRGCLASSESGHVTGSEA-----ELQKVMCMCRSRSRSPRGRGDSAYHSORHSL 1140

QY 162 -PYKPHICTLSVSP---EIDSINS-LTPDNH 187  
Db 1141 TPVTP-----LSPGGLSEEDVNGVMPDTH 1165

Search completed: July 3, 2003, 15:50:45  
Job time : 38 secs

Db 1141 TPVTP-----LSPGGLSEEDVNGVMPDTH 1165

RESULT 15  
US-08-475-352-4  
: Sequence 4, Application US/08475352  
: Patent No. 5916755  
: GENERAL INFORMATION:  
: APPLICANT: Kraus, Matthias H.  
: APPLICANT: Aaronson, Stuart A.  
: TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE  
: TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREOF, AND  
: TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREO  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Suite 400  
: STREET: 133 Carnegie Way, N.W.  
: CITY: Atlanta  
: STATE: Georgia  
: COUNTRY: U.S.A.  
: ZIP: 30303  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/475,352  
: FILING DATE:  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: 07/978,895  
: FILING DATE:  
: APPLICATION NUMBER: US 07/444,406  
: FILING DATE: 01-DEC-1989  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Perryman, David G.  
: REGISTRATION NUMBER: 33,438  
: REFERENCE/DOCKET NUMBER: 1414-028  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (404) 688-9880  
: TELEFAX: (404) 688-9880  
: INFORMATION FOR SEQ ID NO: 4:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1342 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-475-352-4

Query Match 4.6%; Score 81; DB 2; Length 1342;  
Best Local Similarity 24.6%; Pred. No. 13;  
Matches 52; Conservative 26; Mismatches 79; Indels 54; Gaps 12;

QY 8 IMAKRGVAGRKPVTHLLEENKQEFHYTIGPSTPV-TIEPGDRIIVDTKAEFGAINS 67  
Db 978 IKRESGPIAGCPGPHGLTNKKLEEVE--LEPELDLDLDEAEDNLATT-----TLGS 1029

QY 68 EQDIPSOLLKMP-----FLNPONGPTMVNGAEKGVLAIVIESML-----PRGVD--- 112  
Db 1030 ALSLPVGTNNRPGSSLLSPSGYMPMNGNLGSCQ---ESAVSGSSERCPRPVSLHP 1086

QY 113 -PYGICAMIPHFGLTGTLTAMLDPLPEKVRMKLDS-----EKVYWSKRHTL- 161  
Db 1087 MPRGCLASSESGHVTGSEA-----ELQKVMCMCRSRSRSPRGRGDSAYHSORHSL 1140

QY 162 -PYKPHICTLSVSP---EIDSINS-LTPDNH 187  
Db 1141 TPVTP-----LSPGGLSEEDVNGVMPDTH 1165

Search completed: July 3, 2003, 15:50:45  
Job time : 38 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Rat on: July 3, 2003, 15:48:47 : Search time 53 seconds  
(without alignments)  
711.65k Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1748

Sequence: 1 MKWLEESIMAKRGVAGRRKP.....GMVLPKVTGAMLNKKLLV 329

Scoring table: BLOSUM62

Gapop 10.0 : Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 3

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US04\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/RCI\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US04\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US04\_PUBCOMB pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCIOS\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US04\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US63\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US63\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1748	100.0	328	9	US-10-086-082-2
2	151	9.2	416	9	US-10-086-082-2
3	148	8.5	33	9	US-10-086-082-14
4	117	6.7	21	9	US-10-086-082-11
5	101	5.8	525	9	US-10-004-551-16
6	101	5.8	525	9	US-10-004-551-20
7	101	5.8	525	9	US-10-004-551-22
8	101	5.8	526	9	US-10-004-551-24
9	101	5.8	593	9	US-10-004-551-14
10	101	5.8	994	9	US-10-004-551-16
11	100.5	5.7	261	9	US-10-004-551-26
12	100	5.7	23	9	US-10-086-082-12
13	99	5.7	20	9	US-10-086-082-3
14	98	5.6	18	9	US-10-086-082-9
15	96	5.5	1094	9	US-10-086-082-11435
16	94	5.4	19	9	US-10-086-082-10
17	93	5.3	682	10	US-09-798-751-2
18	92.5	5.3	381	9	US-10-128-714-8002
19	91.5	5.2	1848	9	US-09-839-956-6

20	91.5	5.2	1848	9	US-10-080-505-6	Sequence 6, Appli
21	89.5	5.1	635	10	US-09-815-242-11225	Sequence 11225, A
22	87	5.0	850	9	US-09-738-626-5613	Sequence 5613, Ap
23	87	5.0	1907	9	US-09-832-292-39	Sequence 39, Appl
24	86.5	4.9	3338	9	US-10-156-761-8454	Sequence 8454, Ap
25	85.5	4.9	925	9	US-09-738-626-6494	Sequence 6494, Ap
26	84.5	4.8	747	10	US-09-815-242-5077	Sequence 5077, Ap
27	84	4.8	355	9	US-10-128-714-3002	Sequence 3002, Ap
28	83.5	4.8	1083	9	US-09-738-626-6914	Sequence 6914, Ap
29	83	4.7	513	10	US-09-833-745-58	Sequence 58, Appl
30	82.5	4.7	755	9	US-09-738-626-3810	Sequence 3810, Ap
31	82.5	4.7	826	9	US-09-746-660A-92	Sequence 92, Appl
32	82.5	4.7	833	9	US-09-746-660A-90	Sequence 90, Appl
33	82.5	4.7	1221	9	US-09-738-626-5163	Sequence 5163, Ap
34	82.5	4.7	1221	10	US-09-910-891-2	Sequence 2, Appli
35	82.5	4.7	1752	10	US-09-841-132-180	Sequence 180, App
36	82	4.7	640	9	US-10-060-019-30	Sequence 30, Appl
37	81.5	4.7	580	9	US-10-156-761-8637	Sequence 8537, Ap
38	81	4.6	652	10	US-09-815-242-10449	Sequence 10449, A
39	81	4.6	677	10	US-09-815-242-12100	Sequence 12100, A
40	81	4.6	1342	9	US-10-172-620-16	Sequence 16, Appl
41	80.5	4.6	395	9	US-09-738-626-6644	Sequence 6644, Ap
42	80.5	4.6	478	10	US-09-841-132-389	Sequence 389, App
43	80.5	4.6	520	9	US-10-156-761-11162	Sequence 11162, A
44	80	4.5	1504	9	US-10-007-706-1	Sequence 1, Appli
45	79.5	4.5	583	9	US-10-156-761-11791	Sequence 11791, A

#### ALIGNMENTS

RESULT 1  
US-10-086-082-2  
: Sequence 2, Application US/10086082  
: Publication No. US20030687402A1  
: GENERAL INFORMATION:  
: APPLICANT: Riodes, Walter  
: APPLICANT: Naughton, Andrew  
: APPLICANT: Robins, Karen  
: APPLICANT: Shaw, Nicholas  
: APPLICANT: Tinschert, Andreas  
: APPLICANT: Zimmermann, Thomas  
: TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)  
: TITLE OF INVENTION: -3,3,3-TRIFLUORO-2-HYDROXY-2-METHYLPROPIONIC ACID  
: FILE REFERENCE: 32213  
: CURRENT APPLICATION NUMBER: US/10/086,082  
: CURRENT FILING DATE: 2002-02-28  
: PRIOR APPLICATION NUMBER: US/09/214,679  
: PRIOR FILING DATE: 1999-12-30  
: NUMBER OF SEQ ID NOS: 14  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO 2  
: LENGTH: 328  
: TYPE: PPT  
: ORGANISM: Klebsiella oxytoca  
US-10-086-082-2

Query Match	100.0%	Score 1748:	DB 9:	Length 328:
Best Local Similarity	100.0%	Pred. No. 5e-163:		
Matches 328:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY	1	MKWLEESIMAKRGVAGRRKPVTHLTEMOKEFHYTIGPSTVLTVEPCDRIIVDTDA	60	
Dh	1	MKWLEESIMAKRGVAGRRKPVTHLTEMOKEFHYTIGPSTVLTVEPCDRIIVDTDA	60	
QY	61	PEGAINSEQDIPSOILLKMPFINFGNGPIMVNGAKGDVLAIVYTESM:PGVDVPGICAMT	120	
DB	61	PEGAINSEQDIPSOILLKMPFINFGNGPIMVNGAKGDVLAIVYTESM:PGVDVPGICAMT	120	
QY	121	PHFGGLTCTDLTAMLNOPLPKVKRMKIDSEKVKWSKRHTLPYKPHIGTILSVSPEDSIN	180	
DB	121	PHFGGLTCTDLTAMLNOPLPKVKRMKIDSEKVKWSKRHTLPYKPHIGTILSVSPEDSIN	180	

```
QY 181 SLTPDNHGGNMDVDPDTPGSSITYPLVRAPGRLPFGDAHACGAGGEGICGTAFFASITL 240
DB 181 SLTPDNHGGNMDVDPDTPGSSITYPLVRAPGRLPFGDAHACGAGGEGICGTAFFASITL 240
QY 241 KVDLIKNNQLSWPRMNAENIMSGSARPLFADATRTAYRCLLYWLVEDFGFGQWDAVMDL 400
DB 241 KVDLIKNNQLSWPRMNAENIMSGSARPLFADATRTAYRCLLYWLVEDFGFGQWDAVMDL 400
QY 301 SOCGKVRGLGNVDPKPYIVGAMINKNLV 328
DB 301 SOCGKVRGLGNVDPKPYIVGAMINKNLV 328
RESULT 2
US-10-156-761-9482
; Sequence 9482, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/20156.761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272597
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9482
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9482
Query Match
Best Local Similarity 9.2%; Score 161; DB 9; Length 416;
Matches 91; Conservative 42; Mismatches 122; Indels 106; Gaps 18;
QY 44 VLTIFGDRITVDPDRAFF---GAINSEODTPSOLKXKPELFTQNGPDKVNGAKGDIVA 100
DB 32 VAMVKPGAEPMECRDNTDCQVGNKDTANDVRDIDELIPHM--LSGSGVGEAGPGLV 99
QY 101 VVIESMLP-----RGV-----DPYGICAMIPHF 123
DB 90 VDIIDLGVPGQTDGAAGQGWYTGIFAKANGSGFLTDYFFDAYKAWDHRGQAVSRHL 149
QY 124 GGLTGDTLT--AMLNDPLPEKVRMIKLDSEKVVYKSRH-----TIPYKPHGISTLSVSPED 177
DB 150 PGIRFGITGTHPGTAPSAEMLAR-----WTKRQALIDIDIFNR--VFPIAVTP--D 159
QY 178 SINSL-----TPD-----NKGNGVEVDIGPNSITYPLVRAPGRLPFG 217
DB 199 ATNALAGTATGLARRTGEARTVPARENGNEDIKNTRGSAVFYFVHVQAKLGGD 258
QY 218 ARACGGTGEI--CCTAVEFAFATIKVLIKNNQLSW-----PPHNAKPI 261
DB 259 LHFSGDGEITFCG--ATKMGFIDFHLKGGMETYGISTNPFV--FGNVEHYTETLTF 317
QY 262 MSIGSARPLE-----DATR--AYRDLTYWLVE--DFGFEQWCAVMDL--SCGKVRKGNV 312
DB 318 LGISVDHDTDINYLDAT--IAYRACLNAVYFKKFCYSGEQAYLLGSSPLECRSGIV 376
QY 313 C 313
DB 377 D 377
```

```
RESULT 3
US-10-086-082-14
; Sequence 14, Application US/10086082
; Publication No. US29030087402A1
; GENERAL INFORMATION:
; APPLICANT: Brieder, Walter
; APPLICANT: Naughton, Andrew
; APPLICANT: Robins, Karen
; APPLICANT: Shaw, Nicholas
; APPLICANT: Tinschert, Andreas
; APPLICANT: Zimmermann, Thomas
; TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)
; FILE REFERENCE: 3,3,3-TRIFLUORO-2-HYDROXY-2-METHYLPROPIONIC ACID
; CURRENT APPLICATION NUMBER: US/10/086,082
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US/09/214,679
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Klebsiella oxytoca
US-10-086-082-14
Query Match
Best Local Similarity 8.5%; Score 148; DB 9; Length 33;
Matches 29; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 110 GVDPYGICAMIPHFGLGTGDTLAMLNDPLPEK 142
DB 1 GVDPYGIEAMIPHFGLGTGDTLAMLNDQLQPK 33
RESULT 4
US-10-086-082-11
; Sequence 11, Application US/10086082
; Publication No. US20030087402A1
; GENERAL INFORMATION:
; APPLICANT: Brieder, Walter
; APPLICANT: Naughton, Andrew
; APPLICANT: Robins, Karen
; APPLICANT: Shaw, Nicholas
; APPLICANT: Tinschert, Andreas
; APPLICANT: Zimmermann, Thomas
; TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)
; FILE REFERENCE: 3,3,3-TRIFLUORO-2-HYDROXY-2-METHYLPROPIONIC ACID
; CURRENT APPLICATION NUMBER: US/10/086,082
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US/09/214,679
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Klebsiella oxytoca
US-10-086-082-11
Query Match
Best Local Similarity 6.7%; Score 117; DB 9; Length 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 42 EPHYTIGPYSTPVLTIHFCDR 52
DB 1 EPHYTIGPYSTPVLTIHFCDR 21
RESULT 5
US-10-004-551-18
; Sequence 18, Application US/10004551
```





```

: CURRENT APPLICATION NUMBER: US/10/004,551
:
: CURRENT FILING DATE: 2001-2-25
: PRIOR APPLICATION NUMBER: 09/533,949
: PRIOR FILING DATE: 2000-08-10
: NUMBER OF SEQ ID NOS: 110
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 24
: LENGTH: 526
: TYPE: PR1
: ORGANISM: Homo sapiens
: OTHER INFORMATION: n 2033 can be A, G, C,
US-10-004-551-24

```

```

Query Match      5.8%, Score 101, DB 9, Length 525,
Best Local Similarity 23.9%, Prod. No. 0.20, Maps 11
Matches 60; Conservative 28; Mismatches 105; Indels 56; Gaps 11

QY 36 TIGPYSTVLTIERGDRIIVTDRAFEAGINSEQDIPSQLKKFFLNPQNGIMVNGAEK 95
DB 26 TVGKGQAPCISETDGLTAAPTPEQPERGVHVTAP---LKULNHPELEEFLEGLGLEK 93

QY 96 GDVLAVYIESMLPCVDQPYGICAMPHHGGTGLTAMLNDDPLPKVRMIKJDSKXYW 155
DB 84 GD---EURLPALPPQDP-----PAPFTPSLP---RLAKQDSRPFT 129

QY 156 SKRHTL---PYKPHIGTILSVSEIDS--INSLIPDNHGNKNDYDGPQSI--TTPVYRA 208
DB 121 SPTPAMAAYPTQPSKEGWPESFESPMLRITAPLPGPSMAVPTLGPCEIASITPUSRA 180

QY 209 -----DC--CRLEIGD-----NHACGGDGE--ICGTAVFASITTIKY 242
DB 181 WTPTEGGPDGMDGRPWAEEVWSGAGIGTIGTISSTASGDRETTTTTTTTTTTCT 240

QY 243 DLIKNWQLSWP 253
DB 241 PGPCSWNFESGP 251

```

```

RESULT 9
US-10-004-551-14
: Sequence 14, Application US/10004551
: Publication No. US20030004310A1
: GENERAL INFORMATION:
: APPLICANT: SHIMKETS, RICHARD A
: APPLICANT: FERNANDES, ELNA
: TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
: FILE REFERENCE: 15966-559
: CURRENT APPLICATION NUMBER: US/10/004,551
: CURRENT FILING DATE: 2001-12-05
: PRIOR APPLICATION NUMBER: 09/635,949
: PRIOR FILING DATE: 2000-02-10
: NUMBER OF SEQ ID NOS: 110
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 14
: LENGTH: 993
: TYPE: PRI
: ORGANISM: Homo sapiens
: OTHER INFORMATION: n 1755 can be A, G, C, or T
US-10-004-551-14

```

```

Query Match      5.8%  Score 131;  DB 9;  Length 993;
Best Local Similarity 23.9%;  Pred. No. 0.65;
Matches 60;  Conservative 28;  Mismatches 105;  Indels 58;  Gaps 11

QY      36  TIGPVSTPVLTIETPDRIIVYTRDAFEGCA:NSEQDLSQLLKMFLEKPNQCHLVANCAK 95
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DY      26  TVGKGAGPIETDQELAAPIEQPQGVHFVTAPT--LKLNIHPLFEFLQSLER 83
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      96  GDVJAVVISMLPRGVDPVIGICAMIPHEGGITGLDTLAINRCPPEKVRVYKLDSEKVV 155
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DY      84  GD--VELKALPFDQDP-----PAPFTPELP-----RLANQUSPVT 120
      |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      156  SKRHLL---PYKPHIGITLVSPEIDS---INSLTETNHGGNMNDVPEIGPGSG---TVPLDVA 208
      |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

121 SPETAMAAVPTPOPSKFGKFWSPFSESEPMCLRTADLPGEFSAVPTLGPGEIASCTPPSRA 180
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
```

```

Query Match      5.8%; Score 101; DB 9; Length 994;
Best Local Similarity 23.9%; Pred. No. 0.63;
Matches 60; Conservative 28; Mismatches 105; Indels 58; Gaps 111
QY      36  TIGPSIPVLITTEPDRIIVTDRAFEAGAINSEODIPSQIKMPFLMPQNGPIMVNGAEK 95
DB      26  TVKCGAOGIETDGEIIAAPTPEQPERGVHVTAPT--LKLLNHHPLLEEFLQEGLEK 83
QY      96  GDVLAVVIESMPLPRGVDPYGCAMTPHEFGGLTGTDLTAMLNDPLPEKVRNKLKLDSEKVV 155
DB      84  GD-- --EELRPAALPFGQP-----PAPFTPSPLP---RLANDOSRRVET 120
QY      156  SKRHRTL--PYKPHICTISVSEIDS--INSLTPDNHGNGMDVPDIPGSI--TYPLVRA 208
DB      121  SPIPAMAAVPTQPSKEGFWSPESPMRLITAPLPGPSMAVPTLGPGEIASTTPPSRA 166
QY      209  -----PG--GRLLFGD-----AHACOGDGE--ICGTAVEEAS--TTIKV 242
DB      181  WFTQPGPGDMGRPWVAWVSGAGIGTGGITSTASGDDEETTTTIIITITITVQT 240
QY      243  DLIKNWQLSNP 253
DB      241  PGPCSNWFSGP 251

```

```

: RESULT 11
: US-0-004-551-26
: Sequence 26, Application US/10004551
: Publication No. US2003002410A1
: GENERAL INFORMATION
: APPLICANT: SH-MKETS, RICHARD A
: APPLICANT: FERNANDES, FLMA
: TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
: FILE REFERENCE: 15966-559
: CURRENT APPLICATION NUMBER: 75/10/004,551
: CURRENT FILING DATE: 2001-12-25
: PRIOR APPLICATION NUMBER: 09/635,949
: PRIOR FILING DATE: 2000-08-10
: NUMBER OF SEQ ID NOS.:110

```

SOFTWARE: PatentIt: Ver. 2.11  
SEQ ID NO 26  
LENGTH: 261  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-004-551-26

Query Match  
Best Local Similarity 5.7%; Score 100.5; DB 9; Length 261;  
Matches 60; Conservative 31; Mismatches 104; Indels 63; Gaps 12;  
QY 36 TGGPSTVLTPIEDGRIIVTRDAFEGAINSEQDIISQLKMPFLNPONGPIMVNGAEK 95  
DB TGGPSTVLTPIEDGRIIVTRDAFEGAINSEQDIISQLKMPFLNPONGPIMVNGAEK 95  
DB 25 TVGKGACGIEIDGELTAAPTPEKGVHVTAPT--LKLNHGHELEEFLEQSELEK 93  
QY 96 GDLVAVYTESLPGVDVYGICAMLPHEGGLTGIDLTAMLCPLPEKVRMIKLDSEKVVV 155  
DB GDLVAVYTESLPGVDVYGICAMLPHEGGLTGIDLTAMLCPLPEKVRMIKLDSEKVVV 155  
DB 84 GD---EELRFALPGQDP-----PAPTFSPLP---RANGDSRVFT 129  
QY 156 SKRHZL---PYKPHZGTVLSVSEIDS--INSLAPNHGGMVDEIGNGST--TYPIVRA 208  
DB SKRHZL---PYKPHZGTVLSVSEIDS--INSLAPNHGGMVDEIGNGST--TYPIVRA 208  
DB 121 SPTPAAAVPTPOSKSGTSPSESPMLRIIAPLTPGSPMAVPTLPGGTASTTPPSRA 150  
QY 209 -----PQ--GRFVIGD-----AAACCGCGE--LCGIAVEFASITIKV 242  
DB -----PQ--GRFVIGD-----AAACCGCGE--LCGIAVEFASITIKV 242  
DB 181 NTPDGRGDMRGPWAVVVGAGIGIGTGTTSIANGDDFTTCTTCTTCTTCTTCTTCT 240  
QY 243 D-----LKKMQLSWPRM 255  
DB D-----LKKMQLSWPRM 255  
DB 241 PQULPAGLQMKKMGKGL 258

RESULT 12  
US-10-086-082-12  
Sequence 12, Application US/10086082  
Publication No. US20030087402A1  
GENERAL INFORMATION:  
APPLICANT: Brieden, Walter  
APPLICANT: Naughton, Andrew  
APPLICANT: Robins, Karen  
APPLICANT: Shaw, Nicholas  
APPLICANT: Tinschert, Andreas  
APPLICANT: Zimmermann, Thomas  
TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)  
FILE REFERENCE: 32213  
CURRENT APPLICATION NUMBER: US/10/086,082  
PRIOR FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US/09/214,679  
PRIOR FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 23  
TYPE: PRT  
ORGANISM: Klebsiella oxytoca  
US-10-086-082-12

Query Match  
Best Local Similarity 91.3%; Score 100; DB 9; Length 23;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 213 LFTGDAAHACGGCEICGTAVEFA 235  
DB LFTGDAAHACGGCEICGTAVEFA 235

RESULT 13  
US-10-086-082-3  
Sequence 3, Application US/10086082  
Publication No. US20030087402A1  
GENERAL INFORMATION:  
APPLICANT: Brieden, Walter

APPLICANT: Naughton, Andrew  
APPLICANT: Robins, Karen  
APPLICANT: Shaw, Nicholas  
APPLICANT: Tinschert, Andreas  
APPLICANT: Zimmermann, Thomas  
TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)  
FILE REFERENCE: 32213  
CURRENT APPLICATION NUMBER: US/10/086,082  
PRIOR FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US/09/214,679  
PRIOR FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Klebsiella oxytoca  
US-10-086-082-3

Query Match  
Best Local Similarity 5.7%; Score 99; DB 9; Length 20;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MKWLEESIMAKRGVGSARKRP 20  
DB 1 MKWLEESIMAKRGVGSARKRP 20

RESULT 14  
US-10-086-082-9  
Sequence 9, Application US/10086082  
Publication No. US20030387402A1  
GENERAL INFORMATION:  
APPLICANT: Brieden, Walter  
APPLICANT: Naughton, Andrew  
APPLICANT: Robins, Karen  
APPLICANT: Shaw, Nicholas  
APPLICANT: Tinschert, Andreas  
APPLICANT: Zimmermann, Thomas  
TITLE OF INVENTION: METHOD OF PREPARING (S)-OR (R)  
FILE REFERENCE: 32213  
CURRENT APPLICATION NUMBER: US/10/086,082  
PRIOR FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US/09/214,679  
PRIOR FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Klebsiella oxytoca  
US-10-086-082-9

Query Match  
Best Local Similarity 100.0%; Score 98; DB 9; Length 18;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 78 MPFLNPONGPIMVNGAEK 95  
DB 78 MPFLNPONGPIMVNGAEK 95

RESULT 15  
US-10-156-763-11435  
Sequence 11435, Application US/10156761  
Publication No. US20030119028A1  
GENERAL INFORMATION:  
APPLICANT: OKURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI





arrett, R.A.: Ragan, M.A.; Sensen, C.W.; Van der Gast, J.  
submitted to GenBank, April 2001  
A:Description: *Sulfobacillus solfataricus* complete genome.  
A:Reference number: A99139  
A:Accession: B90458  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-318 <KOK>  
A:Cross-references: GB:AE00654; NID:513916160; PIDN:AAK42942.1; GSPDB:GN-C-58  
C:Genetics:  
A:Gene: SSQ2810

Query Match 19.6%; Score 342; DH 2; Length 318;  
Best Local Similarity 35.0%; Pred. No. 74-21;  
Matches 106; Conservative 4%; Mismatches 118; Indels 30; Gaps 11;

Qy 43 PVLTIERGDRIVDTRDAFSAI---NSFOIDPSQLKKMPF--DNFONGPIWNGAKR 97  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 19 PILTISGDITVETKEASNGQVTPSTEKD---LKLDFSKIRHJTGIEVKGAEPGR 74  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 98 VLAVYIESMLPRGVDPVGICAMIPHEGLGTDTLATLNDDPLPEKVMIKLDSEKYVSK 157  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 75 ALVEFELEFRHKG--WGWTGVLPGEFGFLADEOYTAPIDLQGP-AUKIWVD-EKYAYAK 129  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 158 RHIL----PYKHIGTISVS-PEIDINSILTPONHGNNMDVPDIGPGSITYPLVRAPGGR 212  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 130 FGLNLNVKVIYPFPVGVTGTLPPRGKLSITIPPREGNGND;KHLTVCTKLVLPE-VNGAL 189  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 213 LFGDARACOGDEIGCTAVEFAFIITIKYDLTKNNOLSWP-----RMENAEINMSI 254  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 190 LSIGDTHLAOGDEVCOTATEAPMEVTVKRLKNVGLTQPLPVTKVKKEMESDEVIAVP 249  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 265 GSARPLEDATRIAYRDLIYWHVEDQFEQWDAYMLISOCCKVRLGNMVD-PKYTCVGMJN 423  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 250 GIDSNLWNATKAIKGISILSKYMA--PVEAYILASVVVNLRVSEVDVPMNIVTAYLP 307  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 324 KNL 326  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 308 KNV 310  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3  
C72750  
probable acetamidase AP0528 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2003  
C:Accession: C72750  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Takai, D. 1999  
A>Title: Complete genome sequence of an aerobic hyperthermophilic Grenatchaeon, Aeropyrum pernix strain K1  
A:Reference number: A72450; MOLID:99310339; PMID:16982966  
A:Accession: C72750  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-377 <RAW>  
A:Cross-references: DDBJ:AP000059; NID:4510391; PIDN:BAA79495.1; PED:95104179  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: AP0528

Query Match 19.3%; Score 332.5; DP 2; Length 377;  
Best Local Similarity 27.8%; Pred. No. 5.5e-20;  
Matches 99; Conservative 56; Mismatches 120; Indels 81; Gaps 12;

Qy 39 PYSTPVJTIFPGDRIIVDTRDAFEGAINSE---QDIPSQELKMFPNPONGPINWNGAKR 95  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 26 PDIEPAAEVDPGGIVVTETSDALQGSLANPGVDVASADLVN--VHPDTPGVVVRCAK 83  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 96 GDVLAIVIESMLPRGVDPYGICAMIPHEGLGTDTLATLNDDPLPEKVMIKLDSEKYVW 155  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 84 GDLVW---EVLQ-KARPFTVTAIDPGF-----FLRD-LFKPKYKRWIRISGYA 130  
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: E75139  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-298 <RAW>  
A:Cross-references: GB:AJ248285; GB:AL096836; NID:g545867; PIDN:CA849830.1; PID:g545834  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0614

Query Match 17.4%; Score 304.5; DB 2; Length 298;  
Best Local Similarity 30.7%; Pred. No. 9.7e-18;  
Matches 94; Conservative 45; Mismatches 120; Indels 47; Gaps 10;

Qy 35 YTGPGYSTPVLTPGDRIVTDRAFGAINEQDIPSQLKMPF--LNPNQGPIMVNG 92  
Db 14 YSFGPNKKEVARAKGEIV:FOILDALGGQVASED---TIKIDFSRVNATGPLYVEG 70  
Qy 93 AEKGDVLAVYIESMLPRGYDPYIGICAMIPHFGLTGTDTITAMNDPLFEKV-----RMK 147  
Db 71 AKRGGLRDLIDIKVEG-----KGAVVTAPGAGVGGKVEEPOTRICE 114  
Qy 148 LUSEKYVSKRHLPVKPHIGT:SVSPETLSINSLTPNKGMDVPEIGGSGITP:YVR 207  
Db 115 VKDGFVF-KGKIPAMPILGV:GVAUD-EEVPTDTPKSHGGMDINIKKTTTFEVPF 172  
Qy 208 AFGCR:FGDAHACQDGEICGTAVAFASITTKVDLIKMWLSWPRMNAENIMSGAR 267  
Db 173 VDCAV:AIQDLHACVAGCGVCSACVSGEIVRVTPMFG-KLEAFLEIETENSTYV:VSD 211  
Qy 268 KPLEDATRIAYRDLIYWLVEDCFE-----QMD-AYML:SOCKKVALGNMVPKVIY 315  
Db 232 ENLDKATIEAV-----SIGVEALRSNLSMDVAYMGASIVHDSVLSQNDPKIV 262  
Qy 319 GAMLNK 324  
Db 283 NVRIPK 288

RESULT 6  
G72414  
hypothetical protein TM0119 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jun-2000  
C:Accession: G72414  
R:Neeson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Hatt, D.H.; Hickox  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: G72414  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-285 <ARN>  
A:Cross-references: GB:AE001593; GB:AE000512; NID:g4980509; PIDN:AAD35213.1; PID:g494061  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0119

Query Match 17.2%; Score 300.5; DB 2; Length 285;  
Best Local Similarity 28.7%; Pred. No. 1.8e-17;  
Matches 85; Conservative 53; Mismatches 125; Indels 33; Gaps 9;

Qy 35 YTGPGYSTPVLTPGDRIVTDRAFGAINEQDIPSQLKMPF--LNPNQGPIMVNG 92  
Db 11 YSFANMAPVEEYVPGQVFFETDLAGGSYD-----KIDFSKVNATGPGVFNNG 60  
Qy 93 AEKGDVLAVYIESM-LPRGVDYVIGICAMIPHFGLTGTDTITAMNDPLFEKVYMKIKLSE 151  
Db 61 VKPGDTLKVRIKRIELPR-----RGMIVTKGFG-----VIGSEV-EGFTHKEIEE 104

Qy 152 KVVWSKRH---TLPKPHIGILSVSPEDSINSLIPDNHGGNMDVPDIPGSGITYPLVRA 208  
Db 107 K--WAVLEFGVRIP:HPMYGVIGVAP:QEGEYPTGTAHRHGGNMDTKETENVVHLPVFQ 164  
Qy 209 PGGRI:FGDAHACQDGEICGTAVAFASITTKVDLIKMWLSWPRMNAENIMSGAR 268  
Db 155 EGALLALGDVHATMGDGCVCVACEPAKVYVVEIDVSKKE-EIKNPVVTNDAYYIIVSLP 223  
Qy 269 PLEDATRIAYRDLIYWLVEDCFEQMDAYML:SOCKKVALGNMVPKVIYVGAJLNK 324  
Db 224 DIBEALKVETRETWFIQRRKTIPTFDAYMLASLSVDVIGISOLVNPATKARIPK 279

RESULT 7  
AH1894  
hypothetical protein all0706 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AH1894  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irig  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Iabat  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH1894  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-328 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA872663.1; PID:g17130051; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all0706

Query Match 15.4%; Score 269; DB 2; Length 328;  
Best Local Similarity 27.6%; Pred. No. 9.1e-15;  
Matches 91; Conservative 59; Mismatches 140; Indels 40; Gaps 9;

Qy 21 VTUHLTEEMOKERHYTIGPYS---TPVLTEPGDRIVTDRAFGAINEQDIPSQLK 77  
Db 1 MTHILKATKS:VH--GGFSLLEFALKVDSSTIDVETTYG-----YIDKAPERT 54  
Qy 78 MPFLN-----PQ-----NGHVMVNGAKGKGVLAIVYFESMLPRGVDPVGCAMIP 121  
Db 55 PFELDICONLPPEKKTANGPHLITGPIYVKDAPGVDVLEKLEAIAPS--LPVGFNAIRS 112  
Qy 122 HFGGLTG:DLTAMNDPLPEKVRMKL---DSEKYVMSKRHTLPYKPHIGILSVSPETD 177  
Db 113 GWG-----ALPNOFTQPALRFILNLANNTAEFPNPSGKIKIPLTPFFGLVATPEN 164  
Qy 178 SINSLTEPDNKGNNMDVPDIPGSGITYPLVRAFGRI:FGDAHACQDGEICGTAVAFASI 237  
Db 165 ARSSVPYGYGGNLDNRELQAGSRIFLPFVPGCLSGDGHSAQDGEVNVTAIETSMN 224  
Qy 238 TTKVDLIKMWLSWPRMNAENIMSGARPLEDATRIAYRDLIYWLVEDCFEQMDAY 297  
Db 225 GRIQTLTKRDLHTAPIAETPTDITMGFAPTLDAALEQALKNMIDFLERFVNLSPEYAY 284  
Qy 298 MLLSOCKKVALGNMVD-PKYTVGAMLNKLN 326  
Db 285 VLCSLAVNFRITQVNVSPNKGVRHGLPKKL 314

RESULT 8  
AB1776  
hypothetical protein lin2752 [imported] - Listeria innocua (strain Clip11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AB1776  
R:Glaser, P.; Brangeul, L.; Buchricser, C.; Amond, A.; Baquero, F.; Berche, P.; Blo  
L. Dominguez-Serna, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi  
S.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Me  
ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AB1776  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-296 <GLA>  
A:Cross-references: GB:AI592022; PIDN:CA697978.1; PID:gl6415298; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin2752

Query Match 15.0%; Score 261.5; DB 2; Length 299;  
Best Local Similarity 29.0%; Pred. No. 3,4c-14;  
Matches 79; Conservative 44; Mismatches 112; Indels 37; Gaps 7;

QY 43 PVLTIPEGDRIIVDTIRCAFEAGINSQDIPSOQLKMPFLNPQNGPIWNGAEKSDVLAAY 112  
DB 20 PVLTVGDSVVTIKDKHFGQNKELHYGEIDWKQF-SPITGPTICIEARPGMLAVT 78  
QY 103 IESMLPRGVDVPGYICAMIPHFGGLTGDTLAMLNCP-----LPEK-VRMTKLUSEKV 155  
DB 79 IEXIT-----LISKDV-VLLNGHTIGVIDOLLNNCIRVKIKQNK: 119  
QY 154 YWKRHTLPYKPHIGTSLVSPIDSINSL-----TPDNHGNMVDVPIGRGSIITYPIVRAP 209  
DB 120 IYSDE-HVQLQKTIGL:-----KTES-HOTKPPHMTKGGTLDSPDITGATIFLPEVK 175  
QY 210 GGRFLFGDAHACOGDEICGTAVEFASITIKVDLKNWLSFRMENAENISGSASP 269  
DB 176 GAMLHVADIRATGFGKTTSTSAEVAETIRLQLLNKNTAPTPIIHKNNLIGLSELT 235  
QY 270 LEDATRIAYRDLIYVDEFGFQWDAYMLLS 301  
DB 236 TERATKALQNLNVLMSDKITLEDAPFLS 267

RESULT 9  
T39115  
formamide-like protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T39115  
R:Runt, C.; Aves, S.; Mcbougall, R.C.; Rajandream, M.A.; Barrell, H.G.  
submitted to the EMBL data Library, November 1999  
A:Reference number: Z21829  
A:Accession: T39115  
A:Status: preliminary; translated from GB/EMBL/CDHU  
A:Molecule type: DNA  
A:Residues: 1-410 <HUN>  
A:Cross-references: EMBL:AL132779; PIDN:CA660014.1; GSPDB:GN00065; SPDB:SPAC869.04  
A:Experimental source: strain 972h-; cosmid c869  
C:Genetics:  
A:Gene: SPDB:SPAC869.04  
A:Map position: 1

Query Match 13.4%; Score 233.5; DB 2; Length 410;  
Best Local Similarity 26.1%; Pred. No. 1,1c-11;  
Matches 92; Conservative 42; Mismatches 115; Indels 57; Gaps 17;

QY 46 TIFPGDRIIVDTIRCAFEAGINSQDIPSOQLKMPFLNPQNGPIWNGAEKSDVLAAY 101  
DB 34 SIEGETVKECLDWTGCG-KNDSAEITKNVDLTKIHYL---SGPFEKGAEPGSDV 90  
QY 102 YIESMLPRGVDVPGYICAM:PHFGGLTGDTLAMLNCPLEKVRMIKLDSEKIVNSKSH-- 159  
DB 91 EIQDVPLENQPGYSGISFAKENG-----GGFLDEHYKAAKAV-FDEGFCSSRHIP 143  
QY 160 -----TLPIK-----PH-----IGTGS-- 171  
DB 144 GVRFPGLIHFGIIGTAPSKELAEWNRREGVAENPHSTHYWQFLNASTAFAGIAD 203

QY 172 -VSPEIDSINSLT-----PUNHGNMVDVPIGRGSIITYPIVRAPGGKLFIGDAHACQGDGE 226  
DB 204 KESATVASEGARTIFGRPEN-GGNCIDKNLSRGSKVFLPVHVPVHGAKLSIGDLHFSGODGE 262  
QY 227 1--CGTAVEFAS-ETIKVULIKNW-----QLSWPRMENAENISGS 266  
DB 263 ISFCG-AIEAGSITIKKILKNG:SDLAMKSPMYLPGHVEPHFSPSRYLTFEGESVDES 321  
QY 267 APRLEDAIRIAYRDLIYVDEFGFQWDAYMLLSQCKVR--LGNMVD 313  
DB 322 SKQHYLCITTAIKQCLRIYIEYFRREGYNDYOLYLLLS-CAPTOGHVAGIVD 372

RESULT 10  
AC1400  
hypothetical protein lmo2603 [imported] - *Listeria monocytogenes* (strain EGD-e)  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC1400  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,  
J.; Jones, L.M.; Karst, U  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.  
ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1400  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-296 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00681.1; PID:gl6412091; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2603

Query Match 13.3%; Score 233; DB 2; Length 296;  
Best Local Similarity 25.0%; Pred. No. 8e-12;  
Matches 74; Conservative 62; Mismatches 136; Indels 24; Gaps 7;

QY 28 EMOKEFHYTIGPYSTPVLTIPEGDRIIVDTIRCAFEAGINSQDIPSOQLKMPFLNPQNGP 87  
DB 13 EMEKS-----TEPA-RVKDGSVVKTKDHFNGQIHAKQHLHYGEIDWKQF-SPTTGP 63  
QY 88 IMVNGAEKSDVLAAYIESMLPRGVDVPGYICAMIPHFGGLTGDTLAMLNCPLEKVRMIK 147  
DB 64 IYIEEARPGDLLAITIKIELLGTGVFLNG--PNI-GITDILLTS-----NSTRCYK 113  
QY 148 LDSEKIVNSKSHITLPYKPHIGTSLVSPIDSIN-SLTPDNHGNMVDVPIGRGSIITYPLV 206  
DB 114 VENNOITVSEDIHIPRTIGLL---KTEELNPSKVPTKNGGLDSSKITEGATIFLPV 169  
QY 207 RAPGGLRIFGDHACOGDEICGTAVEFASITIKVDLKNWLSFRMENAENISGS 266  
DB 170 EYXGASLHGVNVRATIGFGKITATSAEAPAEVTLPLQIHKRTAPTPIIHYHNLICLAS 229  
QY 267 APRLEDAIRIAYRDLIYVDEFGFQWDAYMLLSQCKVRGNMVDPKYTVGMAL 322  
DB 230 GTTIEKATOKTHNNITLTESDKMTIEDALFLISLQADFQVCKLCKPKNITTSIKL 285

RESULT 11  
T04712  
probable formamide (BC 3.5.1.49) F19F18.40 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
C:Accession: T04712  
R:Barvan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15382  
A:Accession: T04712  
A:Molecule type: DNA

A:Residues: 1-432 <BEV>  
A:Cross-references: EMBL:AL035605  
A:Experimental source: cultivar Columbia; BAC clone F19F18  
C:Genetics:  
A:Map position: 4  
A:Intons: 73/3; 153/1; 196/3; 275/2; 356/3; 392/3  
A:Note: F19F18.40  
C:Keywords: hydrolase

Query Match 11.9%; Score 208; DB 2; Length 432;  
Best Local Similarity 24.1%; Pred. No. 1.6e-09;  
Matches 92; Conservative 44; Mismatches 140; Indels 106; Gaps 16;

Dy 28 EMQKEFHYTIGYSTPVLITISPGDRIVCTDEAFEGAI---NSEQDIPSQLKMPFLNPQ 84  
Dy 20 QODQPLHNRHPEIPPAEYKAGEFFRVEDMAGGVKIDNSASDIKN--LVGITTHL 77  
Dy 85 NGPIWV-----NGAFKGDVLAIVTESMLPRGVDPYGCAMLPHEGGLTGIDITAMINDPL 139  
Dy 78 SGPIRVVDEGVAAKAGDALLAVEICNLGPIPGSEWFGTGSRENG-----SGFLTDR 141  
Dy 140 PEKVMKIDSEKVMYKSRHTPEKP-----HTGLSVSP----- 174  
Dy 132 P-----CATKAIWFFGIYAYSPQIGVRFPGTHPSVGTAFSPNELLRNDRERQ 183  
Dy 175 -EIDSINSL-----TPO-----NKGGMVDPDIFGSIYPIVAPAPG 210  
Dy 144 LEESVSELTCEIEGTTPWERIANEAARTIPRGNGGNCIDIKLSRGSKYLVPVFE 243  
Dy 211 GRLEIGDAHACGGGGEI--CGTAVEFASITTIKVDLIKNNWLSW-----PRMENAEINMS 263  
Dy 244 ANLSGTDMHFSGQGEISFGC-ATEMSGPELKECIFRNGMORYLTPMGTPLHVNPIPE 302  
Dy 264 IGSARE-----LEDAIRAYDCIYWLVE---DRGPTQKAYWLL 300  
Dy 303 IGVVPEFSEWLVFEGISVBSGKHLYDAT-VAYKRAVLNA:GYLFKFGYSKEQVY:LL 362  
Dy 301 SQCG-KVRLGNMVDPKVTGAM 321  
Dy 362 SCCPCEGR:SGIVDSFNVAITL 383

RESULT 12  
S74213  
formamide (EC 3.5.1.49) A - Methylophilus methylotrophus  
C:Species: Methylophilus methylotrophus  
C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: S74213; S78115  
R:Wyborn, N.R.; Mills, J.; Williams, S.G.; Jones, C.W.  
Eur. J. Biochem. 240, 318-322, 1996  
A:Title: Molecular characterisation of formamide from Methylophilus methylotrophus.  
A:Reference number: S74213; MUID:96439058; PMID:8841393  
A:Accession: S74213  
A:Molecule type: DNA  
A:Residues: 1-407 <WBY>  
A:Cross-references: EMBL:X99632; NID:q1480104; PIDM:CAA67953.1; PID:q1480105  
A:Experimental source: strain NCIB 10515  
A:Accession: S78115  
A:Molecule type: protein  
A:Residues: 1-544-47; 172-176; 211-215 <WY>  
C:Genetics:  
A:Gene: fmdA  
C:Keywords: hydrolase

Query Match 11.7%; Score 204.5; DB 2; Length 407;  
Best Local Similarity 25.5%; Pred. No. 3e-09;  
Matches 84; Conservative 38; Mismatches 116; Indels 91; Gaps 14;

Dy 28 EMQKEFHYTIGYSTPVLITISPGDRIVCTDEAFEGAI-----NSEQDIPSQLKMPFL 81  
Dy 15 EQQGLHNHRHPEIPPAEYKAGEFFRVEDMAGGVKIDNSASDIKN--LVGITTHL 72  
Dy 82 NPONGPIWNGAEKGDVLAIVTESMLPRGVDPYGCAMLPHEGGLTGIDITAMINDPL 140

A:Residues: 1-389 <KAW>  
A:Cross-references: GB:AP000004; NID:93236131; PIDM:BA430139.1; PID:dl031082; PID:9;  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by Gen  
C:Genetics:  
A:Gene: PH1041  
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1041

Query Match 10.2%; Score 179; DB 2; Length 389;  
Best Local Similarity 22.5%; Pred. No. 3.7e-07;  
Matches 69; Conservative 32; Mismatches 86; Indels 120; Gaps 11;

Dy 88 IMVNGAEKGDVLAIVTESMLPRGVDPYGCAMLPHEGGLTGIDITAMINDPL 137  
Dy 16 VAVEGAKVGDATATVERIRVLSKASGVDS-----PREGSVGDPVYAKKCPVNE 68  
Dy 138 PIPE-----KVRMTK-----LDSE-----KVW 155  
Dy 69 PWPEFEVWIGEDAIRCKKCGAPVSPFRMNGYTWVFDHEGGVGVTVYDKKIAEMIAKNW 128  
Dy 156 SKRHITLP-----YKPHIGTISVSPESDINSLSLPDWHG----- 188  
Dy 129 -EHHALPKNSKOVPIILIAKADIVCVPSIRPFIQGLGUTIPAVD-----IPDSHNRAGDFG 182  
Dy 189 -----GNMVDPIGPGSIYPLVRAPGGRLFTGDHAHACGGDG 225  
Dy 163 YFLINAPHPYAIKEDYETKLTGDHLVDVSVREGAIVIAVPKVEGGVYAGAAHAEQDG 242  
Dy 226 EICGTAVEFASITTIKVDLIKNNWLSWPRMENAEINMSISARPLEDATRIAYRDLIYWL 285  
Dy 243 EVAGHTIDVTAETIVTEVIKGLNLEGPILLPEEDLP-----PLAKPWKRDEWEKVERL 297  
Dy 286 VEDFGPE 292  
Dy 298 AKKEGLE 304

RESULT 14  
A47696

REF:JLT 14  
A47696







```
Db 15 EDDQGIENRHHPPDLMIAWVKPGDERVECDMDWIGQIIGNDSANDVROW--DLICVHYL 72
QY 82 NPQNGPIMVNGAFKGDVLAIVIESMLPRGVGVYGCAMIIHFQSGIATGTH--TAMLDPLP 110
Db 73 ---SGPIGVEGAEGDGLVWVJLDVGTGIDPSOWS-----FNSLFAKENGSGFLCHFP 122
QY 141 EKVMIKIKDSKVKWSKHHT-----LIPKPHIGCT-----LVSPE- 175
Db 123 EASKTI-WDFHGVTTSRHVPKVKYAGIMHPCLGICLPSKELDIYNNKREGDIIAICPR 181
QY 176 -----DSTNSJTPDNGHGNMDVFDIGSGSTIYPLVRA 209
Db 182 VPPIACPTSGSAYMGRSLSDAARAKAAGARTVPPSDHGGNCDIKNLTKGSRVAFVAV 241
QY 209 PGRSLFICIAHACQDGEI--CGTAVAFASIT--IKVILIKWOLSWPFRMENAIMSIN 246
Db 242 KGGG-SMGDLHFSGGDGEITFCG-AIFMAGYLDIKVSEIK-----DGVKKYKT 244
QY 267 ARPIEDATRE--AYEDLIYK-VEJFGFEQ 293
Db 289 KNPVFOFSPITPTVTD--YMFEGISVDE 315

RESULT 2
AMDA_MYCSM
ID AMDA_MYCSM STANDARD; PRT; 405 AA.
AC Q07838;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetamidase (EC 3.5.1.-).
CS Mycobacterium smegmatis.
CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae.
CC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1772;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-23; 81-100 AND 180-194.
RC STRAIN=NCIC 8159;
RX MEDLINE=93232774; PubMed=8473963;
RA Mahenthalingam E., Draper P., Davis E.O., Colston M.J.:
RT "Cloning and sequencing of the gene which encodes the highly
inducible acetamidase of Mycobacterium smegmatis.";
RL J. Gen. Microbiol. 139:575-583(1993).
CC -!- FUNCTION: ALLOWS ACETAMIDE TO BE USED AS A SOLE CARBON OR NITROGEN
SOURCE.
CC -!- CATALYTIC ACTIVITY: Acetamide + H(2)O - acetate + NH(3).
CC -!- SIMILARITY: STRONG, TO M.METHYLOTROPHUS FORMAMIDASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB EMBL: X57175; CAA04452.1;
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA.1.
KW hydrolase.
FT INIT_MET 0 C POTENTIAL.
SQ SEQUENCE 405 AA; 43933 MW; 05E2E359210EF0F2 CRC64;

Query Match
Best Local Similarity 10.2%; Score 178.5; DB 1; Length 405;
Matches 92; Conservative 39; Mismatches 13; Indels 117; Gaps 19;

QY 46 TIEPGRITVTRDAFEAI-----NSEQDT---PSCLAKMPLFPNGPIMVNGAKG 96
Db 33 TVKPGSEFIECKENTGGIGNNSANDVROVDIAPCHML-----SGPIRVEGSEPS 84
QY 97 DVIANYIESMLP-----RVDPIYGICAMIDFEG----- 124
```

```
Db 85 DALLIVDILD:GPVQTNCPCNGEGMGYSGLFAYKNGGSELTDDYYPDAYKAIWDFHQOCT 144
QY 125 -----GLTGI-----DLTAMLDNLPPEKVRMIKIDSEKVVYWSKRHTLPYKP 165
Db 145 SRHVPGVAYTITIHPLGLFTAPSPDLAKWNE---RERAA--LATDPDRV-----PPLALPP 196
QY 166 HL-GFL--SVSPET-----DSINSLTPDNRHGNMDVFDIGSGS--ITYPLVPRAPGGKLF 215
Db 197 LDVDTLGTAGSDLLQAIANDGARTVPPRENGGHHDIKNFTRSRTPYP-VFVEGAMLSG 255
QY 276 GDIAHACQDGEI--CGTAVAFASIT--IKVILIKWOLSWPFRMENAIMSISGARPL--- 270
Db 256 GDLHFSQDGEINFCG-AIEMGGFIDMEVDLIKGMETVGT-NP--IFMGRVPELYSE 312
QY 271 -----EGATRIAYEDLIYKVE---DFGFEQWDAYMLL 300
Db 313 WLTFIGISVDRHAENRNAYMDAT-MAYRNACLNATLEYLKKWGYTCEQAYIL 362

RESULT 3
AMDA_MYCSM
ID AMDA_MYCSM STANDARD; PRT; 950 AA.
AC Q5C968;
DT 31-NOV-1997 (Rel. 35, Created)
DT 31-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exonuclease ABC subunit A.
GN UVRA.
OS Neisseria gonorrhoeae.
CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97340932; PubMed=9197406;
RA Black S.G., Fyfe J.A.M., Davies J.K.:
RT "Cloning, nucleotide sequence and transcriptional analysis of the
uvrA gene from Neisseria gonorrhoeae.";
RL Mol. Gen. Genet. 254:479-485(1997).
CC -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVRA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB FMBL: 034760; AAA8485.1;
DR InterPro: IPR003439; ABC_transportr.
DR LateralPro: LPR004602; UvrA.
DR Pfam: PF00005; ABC_tran; 2.
DR ProDom: PD000006; ABC_transportr; 1.
DR TIGRFAMs: TIGR00630; uvrA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW DNA-binding; Zinc-finger.
FT NP_BIND 42 49 ATP (POTENTIAL).
FT NP_BIND 649 656 ATP (POTENTIAL).
FT ZN_FING 262 289 C4-TYPE.
FT ZN_FING 748 774 C4-TYPE.
SC SEQUENCE 950 AA; 105552 MW; 359806240AADD6F1 CRC64;

Query Match
Best Local Similarity 5.7%; Score 99.5; DB 1; Length 950;
Matches 80; Conservative 56; Mismatches 127; Indels 125; Gaps 20;
```

```
QY 12 RGVGAGRPVTHLTEMQKEFHVITIGPYSTPVLTIIEPGDRIIVDTDAFEGAINSEDDI 71
DB 544 RDLGNSVIVVEHDEDA:READFVVDMPGAG-----EHGNNVLI--ADTPENVAKCENS 596
QY 72 PSQLI-----KMPFLNPONGPIWNGAEKGDVLAIVTESMLPGVDVPYGCAMI 120
DB 597 TCQYLGKSKSTAVPSERTP-VNP-GRMLVLKGAGNNLKNVTIE--LPLG-----LI 644
QY 121 PHFGGLTGED:TAMINDPLPEKVRMIKDSKVKYWSKRHILPYKPHLGLTSLVSFEIDSIN 180
DB 645 TCITGVSGSGKSTLINDTL-----AKITARELNRAQEEAPY-----DDIR 685
QY 181 SLTPDNHGGNMVDPJG-----EGSIT-----YPLVRAPG---GRIFI--- 215
DB 646 GLEHLDKVINDQSPICRTPKRNPATVTLGTPITRELFAQVPLSRGSGYVWGRFSENVKG 745
QY 216 GDAHACQDGG-----EIC-----GTAVE--- 233
DB 746 GRCEACQDGGVIVKEMHFLPDVYVPCVCHCKRYNRETLEIQYKGNISQVLDMTVEEAC 805
QY 234 --FASITTI--KVDLIKNNWLSWPMENAEINISGARPLEDAIRAYRD--LYWLV 286
DB 806 EFFDAVPIVSRKLOT:MDVGLGYIRIGQSATTLSSGGEAQVRKVLAEULSKRDTGYILD 865
QY 287 E-DFGEQWMDAYMLLSQCGKVR-IGNMV 312
DB 866 EPTTGLHFAIDIA:LEVICR-KGKGNST 893
RESULT 4
UVRA_NEIMA
ID UVRA_NEIMA STANDARD: PRT: 948 AA.
AC Q4JUS4:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exonuclease ABC subunit A.
GN UVRA OR NMA1159.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=656599;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=15761913;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown G., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skellern J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria
RA meningitidis 22491";
RA Nature 404:502-506(2000).
CC -!- FUNCTION: The ABC excision nuclease is a DNA repair enzyme that
CC catalyzes the excision reaction of UV-damaged nucleotide segments
CC producing oligomers having the modified base(s). UVRA is an ATPase
CC and a DNA-binding protein that preferentially binds single-
CC stranded or UV-irradiated double-stranded DNA (By similarity).
CC -!- SUBUNIT: Consists of three subunits; uvrA, uvrB and uvrC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL162755; CAB84421.1; .
```

```
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR004602; UvrA.
DR Pfam: PF00005; ABC_tran; 2.
DR Pfam: PF00006; ABC_transportr; 1.
DR TIGRfams: TIGR00630; uvrA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW DNA-binding; Zinc-finger; Complete proteome.
FT NP_BIND 42 49 ATP (POTENTIAL).
FT NP_BIND 649 656 ATP (POTENTIAL).
FT ZN_FING 262 289 C4-TYPE.
FT ZN_FING 748 774 C4-TYPE.
SQ SEQUENCE 948 AA; 105191 MW; E58717DF762B9A0D CRC64;
Query Match 5.6%; Score 98.5; DB 1; Length 948;
Resid Local Similarity 20.6%; Pred. No. 2.9;
Matches 80; Conservative 55; Mismatches 128; Indels 125; Gaps 20;
QY 12 RGVGAGRPVTHLTEMQKEFHVITIGPYSTPVLTIIEPGDRIIVDTDAFEGAINSEDDI 71
DB 544 RDLGNSVIVVEHDEDA:READFVVDMPGAG-----EHGNNVLI--ADTPENVAKCENS 596
QY 72 PSQLI-----KMPFLNPONGPIWNGAEKGDVLAIVTESMLPGVDVPYGCAMI 120
DB 597 TCQYLGKSKSTAVPSERTP-VNP-GRMLVLKGAGNNLKNVTIE--LPLG-----LI 644
QY 121 PHFGGLTGED:TAMINDPLPEKVRMIKDSKVKYWSKRHILPYKPHLGLTSLVSFEIDSIN 180
DB 645 TCITGVSGSGKSTLINDTL-----AKITARELNRAQEEAPY-----DDIR 685
QY 181 SLTPDNHGGNMVDPJG-----EGSIT-----YPLVRAPG---GRIFI--- 215
DB 646 GLEHLDKVINDQSPICRTPKRNPATVTLGTPITRELFAQVPLSRGSGYVWGRFSENVKG 745
QY 216 GDAHACQDGG-----EIC-----GTAVE--- 233
DB 746 GRCEACQDGGVIVKEMHFLPDVYVPCVCHCKRYNRETLEIQYKGNISQVLDMTVEEAC 805
QY 234 --FASITTI--KVDLIKNNWLSWPMENAEINISGARPLEDAIRAYRD--LYWLV 286
DB 806 EFFDAVPIVSRKLOT:MDVGLGYIRIGQSATTLSSGGEAQVRKVLAEULSKRDTGYILD 865
QY 287 E-DFGEQWMDAYMLLSQCGKVR-IGNMV 312
DB 866 EPTTGLHFAIDIA:LEVICR-KGKGNST 893
RESULT 5
AMID_SULSO
ID AMID_SULSO STANDARD: PRT: 504 AA.
AC P55896;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Amidase (EC 3.5.1.4).
GN SSO2122 OR C02016 OR C02_017.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX STRAIN=ATCC 49255 / DSM 5833 / MT-4;
RX MEDLINE=21346015; PubMed=11453462;
RA d'Abusco A.S., Amendola S., Scandurra R., Politi L.;
RT "Molecular and biochemical characterization of the recombinant amidase
RT from hyperthermophilic archaeon Sulfolobus solfataricus.";
RL Extremophiles 5:183-192(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=97055432; PubMed=8899719;
RA Sørensen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.-Y.,
```

```

RA Liu Q.Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
RA Docliffie W.F., Ragan M.A., Charlebois P.,
RT "Organizational characteristics and information content of an
RT archaeal genome: 156 kb of sequence from Sulfolobus solfataricus
RT P2.",
RT Mol. Microbiol. 22:175-191(1996).
RN (3).
RN SEQUENCE FROM N.A.
RC STRAIN:ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=1142726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awaye M.J., Chan-Welher C.C.Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Pena X.,
RA Charlebois R.B., Docliffie W.F., Duguet M., Gaasterland T.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.,
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.",
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -!- FUNCTION: Enantioselective, active on 2-to 6-carbon aliphatic
CC amides and on many aromatic amides
CC -!- CATALYTIC ACTIVITY: A monocarboxylic acid amide + H(2)O -> a
CC monocarboxylate + NH(3).
CC -!- SUBUNIT: Monomer.
CC -!- MISCELLANEOUS: Active over the pH range 4-9 and at temperatures
CC from 60 degrees to 95 degrees Celsius.
CC -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CR EMBL: AF260611; AAK83092.1; -
DR EMBL: Y08256; CAA69448.1; -
DR EMBL: AE006819; AAK42301.1; -
DR InterPro: IPR006120; Amidase.
DR Pfam: Pf01425; Amidase; 1.
DR PROSITE: PS00571; AMIDASES; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 504 AA; 55655 MW; A9103AB7D09A88EE CRC64;

Query Match 5.6%; Score 97.5; DB 1; Length 504;
Best Local Similarity 21.4%; Pred. No. 1.5;
Matches 54; Conservative 32; Mismatches 29; Indels 77; Gaps 11;

QY 79 PFLPNQCPIMVNGAEKGVAVIESMLPRGVDPYGCICAMIPFGGLIGTDLTLMNDP 138
DB 157 PVLPNPNPEYVAGSSGSAVAV-----ASGYCDMA--VGGDQGS----- 135
QY 139 LPKVRMIKLDSEKV-YMSKHTLPYKPHIGTILSVSPFIDSN-----SLTPQNHG 188
DB 196 -----TRIPSSVWGIVGLKPTKGLVPTGAFSTETLRLGPMANTVKDVALLEVA 248
QY 189 CNMVDVPLGCPISYPLVRAAGGRIFGDAHACGCEIGCATVEFASITIKVKLIKNN 248
DB 249 GRDELSDQPSLPPPVYK-PYSKLIDG-----VSDMKVGIYKE- 287
QY 249 QLSWPRMENAMINISGARPLEDAIRIAYRLIYML-VED-----FGFQGMAYML 303
DB 288 GFNNSENKDVLDYLSAKKLED-----YGKVEDTSTPLRMGLDITPFAIE 337
QY 301 SQCGKVRIGNMV 312
DB 338 CATATMILGSGV 349

RESULT 6
UVRA_NEIMB STANDARD: PRT: 949 AA.
ID UVRA_NEIMB

```

```

AC 2942P1:
CT 15-JUN-2002 (Rel. 41, Created)
LT 15-JUN-2002 (Rel. 41, Last sequence update)
RT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exonuclease ABC subunit A.
GN UVRA OR NMB0962.
GS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN (1).
RN SEQUENCE FROM N.A.
RX STRAIN:MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Kachin K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.",
RL Science 287:1809-1815(2003).
CC -!- FUNCTION: The ABC excision nuclease is a DNA repair enzyme that
CC catalyzes the excision reaction of UV-damaged nucleotide segments
CC producing oligomers having the modified base(s). UVRA is an ATPase
CC and a DNA-binding protein that preferentially binds single-
CC stranded or UV-irradiated double-stranded DNA (By similarity).
CC -!- SUBUNIT: Consists of three subunits: uvraA, uvrB and uvrC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. UVRA SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002447; AAF41368.1; -
DR TIGR: NMB0962; -
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR004602; UVRA.
DR Pfam: PF00005; ABC_tran; 2.
DR ProDom: PD000006; ABC_transportr; 1.
DR TIGRFAMs: TIGR00630; uvra; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW DNA-binding; zinc-finger; Complete proteome.
FT NP_BIND 42 49 AIP (POTENTIAL).
FT NP_BIND 649 656 AIP (POTENTIAL).
FT ZN_FING 262 285 C4-TYPE.
FT ZN_FING 748 774 C4-TYPE.
SQ SEQUENCE 949 AA; 105536 MW; 153F5AE2D48CDD35 CRC64;

Query Match 5.6%; Score 97.5; DB 1; Length 949;
Best Local Similarity 20.4%; Pred. No. 3.5;
Matches 75; Conservative 57; Mismatches 127; Indels 125; Gaps 20;

QY 12 RGVGAGRPVTHLTEMCKEFHTTIGYTPVLTIEPGRIIVDTRDAPEGAINSQDI 71
DB 544 RDLGNSVIVVEHDEUAIREADQVVDVMDGPGAG-----EHGGNVLII--ADTENVACNSV 596
QY 72 PSQLL-----KMPFLNPONGPIVNGAEKGVAVIESMLPRGVDPYGCICAM 120
DB 547 TGOYLSGKKSIAPVSEKTP-VNPDRLVVLKAGRNKLVNLTLE--LPLG-----LI 644
QY 121 PFGGIGTGLTAMNDPLPEKVRMIKLDSEKVYMSKHTLPYKPHIGTILSVSPFIDSN 180
DB 645 TCITGVSGSGKSLINDTL-----AKITARELNRAQEEPAFF-----DDIH 685

```

```
QY 181 SLTDNNGGNDVDPIC-----PGSIT-----YFLVAFG---SKLF1 --- 215
DB 686 GLEHLKVINVDOSPGRTPRSNPNATYGLFTTIRLEFAGVILSRGNGVNGWSPKFKVAG 745
QY 216 GDAHACOGD-----EIC-----GTAVE--- 233
DB 746 GRACOGGCV:KVENHFLPDVYVPOVCHGKRYNRETLETKYGNISQVLMVFEAR 805
QY 234 --FASITTI--KYDLINKWOLSWPRMENAINISGARPLEDATRIAYRDLYLWVDEGFGQWDA 296
DB 806 EFFDAVTVSRKLGJLMDVCGVIRLGOSATTILSSGCAQVRLKALELSKRDGRTGLYLLD 865
QY 287 E-DGFFQWDAYMLLSOGKVR-LGNMV 312
DB 866 EPTTGLHFADIALLELVIGRLKKGKNSI 893

RESULT 7
GALT_THETN
AC Q8R86:
STANDARD: PRT: 519 AA.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Galactose-1-phosphate uridylyltransferase (EC 2.7.7.10) (UniProt)
DE uridylyltransferase.
GN GALT OR TTE1929.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M84 / JCM 11007;
RX MEDLINE=21592816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:889-700(2002);
CC -!- CATALYTIC ACTIVITY: UDP + alpha-D-galactose 1-phosphate =
CC diphosphate + UDP-galactose.
CC -!- PATHWAY: Galactose metabolism; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (potentially).
CC -!- SIMILARITY: BELONGS TO THE GALACTOSE-1-PHOSPHATE
CC URIDYLTRANSFERASE FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL: AEO13143; AN25108.1;
CC DR PROSITE: PS01163; GAL_F_UDP_TRANSFER_II; FALSE_NEG.
KW Transferase; Nucleotidyltransferase; Galactose metabolism;
KW Complete proteome.
SQ SEQUENCE 519 AA: 60467 MW: 60467 MW: 60467 MW: 60467 MW: 60467 MW:
Query Match 5.48; score 94.5; DB 1; length 519;
Best Local Similarity 20.28; Pred. No. 2.8;
Matches 56; Conservative 42; Mismatches 94; Indels 85; Gaps 13;

QY 83 PONGPIMVNGAKGDVLAVIESMLPRGVDPYGLICAMIPHGSLGTGLIAMLNWLPK 142
DB 158 PEKDPKEIAKAKK-----IPQ--SGYFKLGLCIENVGFGN-----LNHPARON 199
QY 143 VRMIKLD-----SEKVVSKRHTLPYKPHUGITLSVSP-----IDSKSLTPDNH 187
DB 200 LRITPVKVGQWYFQSPVYVYNEHCILLHEHI-PMKISEKTFVRLDFIDDFHYFM 258
```

```
QY 188 GGNMDVDPIDGSGI-----TYPLVKAHGGRIFIGDAHACVSDGDCIGCIGIAVEFAS 236
DB 259 GSNADLPIVG--GSILVHFHFGQGRHTPEEAEPTVEYFHPKY----- 360
QY 237 IITTKVLLINKWOLSWPRMENAINISGARPLEDATRIAYRDLYLWVDEGFGQWDA 296
DB 301 -KEYKAGTLK-WPNSVIRI-----SSKOREKLTLSSHILNWN--KGYSDSDVDV 346
QY 297 YMLLSOGKVRLGNMVDP-----KYTVGAMLNKN 325
DB 347 LAYSEKDGKIVPHNTITPIARENKGEYEYIDLVLNN 363

RESULT 8
NIA_ASPNG
TD NIA_ASPNG STANDARD: PRT: 867 AA.
AC P36858:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Nitrate reductase [NADPH] (EC 1.6.6.3) (NR).
GN NIAD.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92175518; PubMed=1541396;
RA Jaxles S.E., Campbell E.L., Punt P.J., Hawker K.L., Contreras R.,
RA Hawkins A.R., van den Hondel C.A.M.J.J., Kinghorn J.R.;
RT "The Aspergillus niger niaD gene encoding nitrate reductase: upstream
RT nucleotide and amino acid sequence comparisons.";
RL Gene 111:149-155(1992);
CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -!- CATALYTIC ACTIVITY: NADPH + nitrate = NADP(+) + nitrite + H(2)O.
CC -!- COFACTOR: REQUIRES FAD, A HEME GROUP (CALLED CYTOCHROME B-557)
CC AND ONE MOLYBDENUM ATOM.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
CC N-TERMINAL DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL: M77022; -; NOT_ANNOTATED_CDS.
CC DR PIR: JQ1525; JQ1525.
CC DR HSP: P17571; 2CND.
CC DR InterPro: IPR001199; Cyt_B5.
CC DR InterPro: IPR001834; Cyt_B5_reductase.
CC DR InterPro: IPR000572; Euk_Mb_oxred.
CC DR InterPro: IPR005066; Mo-co_dimer.
CC DR InterPro: IPR001433; Oxred_FAD/NAD(P).
CC DR Pfam: PF00173; heme_1; 1.
CC DR Pfam: PF00174; oxidored_molyb; 1.
CC DR Pfam: PF00970; FAD_binding_6; 1.
CC DR Pfam: PF03404; Mo-co_dimer; 1.
CC DR PRINTS: PR00406; CytB5RDASE.
CC DR PRINTS: PR00363; CYTOCHROME_B5.
CC DR PRINTS: PR00407; EUMOPTERIN.
CC DR ProDom: PD000612; Cyt_B5; 1.
CC DR PROSITE: PS00291; CYTOCHROME_B5_1; 1.
CC DR PROSITE: PS50255; CYTOCHROME_B5_2; 1.
```



```

SQ SEQUENCE 506 AA: 57060 MW: 7397FD52A4987656 CRC64:
Query Match: 5.28: Score 90.5; DB 1: Leuqth 50C;
Best Local Similarity 25.98: Pred. No. 5.7;
Matches 52: Conservative 32; Mismatches 74; Indels 43; Gaps 13;

QY 14: EKVRMVKLSE-----KVYKSKRHITLDPYKPHGTLSVSHFIDSTNSILTPDHHGGMND 152
DB 2 KKAKMI-VDKYKIGEVDRKIYS-----FTFHKGR-AVVEGIVEPDHPDEDEGFRKD 53
QY 193 VPDIGPGSITVPLVAPGGRLFTGDARACQGGEGCG--TAVEFASITTIKVDLIKNMQL 250
DB 54 VQSL-IKELQVPIRYDGGN-FLSGYMWELVGVENRPRDLIAWQTITENVGITNEFL 111
QY 251 SWPMKNENNTMSIG-SARPLEDATRIAYRDLI-----YK--IVEDFGFQWDAYM 298
DB 112 SWPKKVAITEVNMAYNLTOTRID-----AARNLVEYCNHPKGYNSDLRSRSHGYEQ--PYG 164
QY 299 LLSOCGKVRKGNMVDPKYTVG 319
DB 165 IKTWC----LGNEMDGPWQIG 181

RESULT 11
FTH1:HAETN
ID FTH1:HAETN STANDARD; PRT: 635 AA.
AC P71377;
DT 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell division protein ftsh homolog 1 (EC 3.4.24.-).
GN FTSH-A OR H1335.
OS Haemophilus influenzae.
CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
CC Haemophilus
CC Haemophilus
OX NCBI_taxid=727;
RN 11
SEQUENCE FROM N.A.
SI:RAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95150630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott C.D., Shirley R., Liu L.T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gachin C.L., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RA Rd.";
RI Science 269:496-512(1995).
RE 21
REVIEWS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.;
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE
CC (BY SIMILARITY).
CC 1- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).
CC 1- SUBCELLULAR LOCATION: Integral membrane protein.
CC 1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32812; AAC22979.1;
CC TIGR: H1335;

```



```
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003599; AAA_ATPase_centr.
DR InterPro: IPR003960; AAA_sub.
DR InterPro: IPR000642; Peptidase_M41.
DR Pfam: PF00004; AAA; 1.
DR TIGRfam: TIGR01241; PtsH_fam; 1.
DR SMART: SMC0382; AAA; 1.
DR PROSITE: PS00674; AAA; 1.
KW Zinc; Complete proteome.
KW Zinc; Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 24 POTENTIAL.
FT DOMAIN 25 95 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 96 120 POTENTIAL.
FT DOMAIN 121 635 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 191 198 ATP (POTENTIAL).
FT METAL 413 413 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 414 414 BY SIMILARITY.
FT METAL 417 417 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 635 AA; 73030 MW; 70E6DD1E295839D CRC64;

Query Match 5.18; Score 89.5; DB 1; Length 635;
Best Local Similarity 20.48; Proc. No. 9.4;
Matches 80; Conservative 59; Mismatches 117; Indels 143; Gaps 23;

QY 13 GVAGAKPVTHLITEEMKEFHVTIGPYSTFVITIEGDMHIVLTQAF-----ES 63
Db 226 GVASR---VRDPEQAKK-----NAPCI-----IFIDFANVHRCAGLGG 266
CY 64 AINSEDIPLSKMPFLNFGNPIKVNKAKEGV----- 98
Db 267 HDREQTLNCLVMDEGFSNGGVIVIAATNREFVLPALTRGFEURQVYVGLFVKKGR 326
QY 99 ---LAVVIESM-LPFGVDPGICAMPHFGCLGIDLTAMLNPL----- 139
Db 327 EQILKVMKRVSAQVDVAMTLARGTP---GYSGADLANLVAAALFAVRNKRVTINE 183
QY 140 -----PEKVRMIKLDSEK---VYWSKRHILPKPHICITLSVSPEDIS:NSLIP 185
Db 384 FEAKOKINMGPERKTMTOKKSTAYHEAGHAI-----VGYL--VPESDPVKVT-- 434
QY 186 NHGGMNDVDPGISTYPLVRAGGRFLFGDAHACGGGEICGTAVEFASITTIKVDLI 245
Db 435 -----IPRGRALGVTFPLP-----EGD-QTISOKQLES-----KISTL 468
QY 246 KNWLSNPRMENAIMSGSARPLEDATRIAYRDLIYVLVDFGP-EQWDATMLLSQG 304
Db 469 YAGRLAEDLIYGENI-STGASNDIKVATNIARNMYTQW-----GFSEKLGPILYTEDEG 522
QY 305 KVRIG-NMVPKY-----TVGAMLNKN 325
Db 523 EVELGRSKAKAKHMSDETAHSIDEVRAIVNRN 553

RESULT 12
PK_ECOLI
ID PK_ECOLI STANDARD: PRT: 687 AA.
AC P28688; Q47549;
DT 01-DEC-1992 (Rel. 24, Created)
DI 01-DEC-1992 (Rel. 24, Last sequence update)
DI 15-JUN-2002 (Rel. 41, Last annotation update)
DE Polyphosphate kinase (EC 2.7.4.1) (Polyphosphoric acid kinase) (ATP-
DE polyphosphate phosphotransferase).
GN PPK OR B2501 OR Z3764 OR ECS3363.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
[1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-K12;

MEDLINE-93054553; PubMed-1331061;
AKiyama M., Crooke E., Kornberg A.;
RT "The polyphosphate kinase gene of Escherichia coli. Isolation and
RT sequence of the ppk gene and membrane location of the protein.";
KL J. Biol. Chem. 267:22556-22561(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / XG1655;
RC MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.E., Rode C.K., Mayhew G.F.,
RA Gregor C., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RC MEDLINE-97345980; PubMed-9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isoro K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizutuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sempel G., Satch Y., Sivasubram S.,
RA Tagari H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yaraagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:191-111(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL53 / AFCC 700927;
RC MEDLINE-21074935; PubMed-11206552;
RA Porcia N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Postbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharam T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMC 0509952;
RC MEDLINE-21156231; PubMed-11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 1-90 FROM N.A.
RC STRAIN-K12;
RC MEDLINE-87260114; PubMed-3301938;
RA Smith J.M., Daum H.A. III;
RT "Identification and nucleotide sequence of a gene encoding 5'-
RT phosphoribosylglycinamide transformylase in Escherichia coli K12.";
RL J. Biol. Chem. 262:10565-10569(1987).
RN [7]
RP SEQUENCE OF 678-687 FROM N.A.
RC STRAIN-K12;
RC MEDLINE-93107072; PubMed-8380170;
RA Akiyama M., Crooke E., Kornberg A.;
RT "An exopolyphosphatase of Escherichia coli. The enzyme and its ppx
RT gene in a polyphosphate operon.";
RL J. Biol. Chem. 268:633-639(1993).
RN [8]
RP SEQUENCE OF 433-435, AND MUTAGENESIS OF HISTIDINE RESIDUES.
RC MEDLINE-9721395; PubMed-8962061;
RA Kumbic K.D., Ann K., Kornberg A.;
RT "Phosphohistidyl active sites in polyphosphate kinase of Escherichia
```

```

RT  coli."
RL  Proc. Natl. Acad. Sci. U.S.A. 93:14391-14395(1996).
CC  -!- FUNCTION: Catalyzes the reversible transfer of the terminal
CC  phosphate of ATP to form a long-chain polyphosphate (polyP).
CC  Can form linear polymers of orthophosphate with chain lengths up
CC  to 1000 or more. Can also act in the reverse direction to form
CC  ATP in the presence of excess ADP. Can also use GTP instead of
CC  ATP, but the efficiency of GTP is 5% that of ATP.
CC  -!- CATALYTIC ACTIVITY: ATP + (phosphate)(N) = ADP + (phosphate)(N+1).
CC  -!- SUBUNIT: HOMOTETRAMER.
CC  -!- SUBCELLULAR LOCATION: Inner membrane-associated
CC  -!- PTM: An intermediate of this reaction is the autophosphorylated
CC  ppk in which a phosphate is covalently linked to histidine
CC  residues through a N-P bond.
CC  -!- SIMILARITY: BELONGS TO THE POLYPHOSPHATE KINASE FAMILY.
CC  -!- SIMILARITY: CONTAINS 1 PLD PHOSPHOTRANSFERASE DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL: L03719; NOT ANNOTATED_CDS.
DR  EMBL: AC000336; AAC75554.1;
DR  EMBL: D9878; BAA16389.1;
DR  EMBL: AP005479; AAG57611.1;
DR  EMBL: APC02551; BAB36786.1;
DR  EMBL: M13747; AAA03903.1; ALT.INIT.
DR  EMBL: L06129; NOT ANNOTATED_CDS.
DR  F01: B28486; B28486.
DR  F01: A44306; A44306.
DR  EcoGene: EG1510; ppk.
DR  InterPro: IPR001736; PJD.
DR  InterPro: IPR003414; PP_kinase.
DR  Pfam: PF02503; PP_kinase; 1.
DR  PROSITE: PS50035; PLD; 1.
KW  Transferase; Phosphorylation; Inner membrane; Complete proteome.
FT  INIT_MET 0 0
FT  DOMAIN 429 403 PLD PHOSPHODIESTERASE.
FT  ACT_SITE 434 434 FORMS THE PHOSPHOTRANSFERASE INTERMEDIATE.
FT  ACT_SITE 453 453 FORMS THE PHOSPHOTRANSFERASE INTERMEDIATE.
SQ  SEQUENCE 687 AA; 80300 MW; D040A2AD3D726804 CRC64;

Query Match 5.18; Score 89.5; DB 1; Length: 667;
Best Local Similarity 23.5%; Pred. No. 10;
Matches 47; Conservative 39; Mismatches 69; Indels 45; Gaps 12;

QY 25 LTRKKQKEPHYTYGYPSTPV---LTIFPGD---RIIVDTIRDAFEGALNSEQIIPSGE-LK 77
DB 479 ITNEVRVFNFIENPY-RPVTEFLMVSPONSRELLIEMVD--REIANAQGLFSGITLK 535

QY 78 MFLPLNPQGPIMVNGAEKGVAVIESM--LPRGVDPYGGICAMTPHFGGLTGDTLAML 135
DB 536 ---LN-----NLVDKGLVDRDLAASSGSPVNLVSGMCS:SPNLEGISDNIRAI 564

QY 136 NDLPLPKVMKIDS---EKVY-----NSKRFLPLVYKHGLTSLVSPET----- 176
DB 585 VDRYLEHDRVYTFENGSDKKVLSNADWMTNR-IDYREVAVPLDPLKORVLTIDIL 643

QY 177 --DSINSLTPDNHGGNMVDP 194
DB 644 FSDTVARVYDKLELSNRYVF 653

RESULT 13
ID IGA4_HAEIN STANDARD; PRT: 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

```

```

DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE  Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN  IGA.
OS  Haemophilus influenzae.
CC  Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC  Haemophilus.
CC  OX  NCBI_TaxID=727;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=NHII HK61;
RX  MEDLINE=92234949; PubMed=1373717;
RT  "A comparative genetic study of serologically distinct Haemophilus
RT  influenzae type 1 immunoglobulin A1 proteases.";
RC  J. Bacteriol. 174:2913-2921(1992).
CC  -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC  PRODUCING INTACT FC AND FAB FRAGMENTS.
CC  -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC  certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC  substrates are known.
CC  -!- SUBCELLULAR LOCATION: Secreted.
CC  -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC  SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC  OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC  DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC  SIMILARITY). BELONGS TO PEPTIDASE FAMILY S6.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL: M87491; AAA24968.1;
DR  MEMOPS: S05.001;
DR  InterPro: IPR000710; Iga_S6.
DR  InterPro: IPR004899; Peract_sup.
DR  Pfam: PF02395; IGA1; 1.
DR  Pfam: PF03212; Peractin; 2.
DR  PRINTS: PR00921; IGASERPTASE.
KW  Hydrolyase; Serine protease; Transmembrane; Signal.
FT  SIGNAL 1 25 POTENTIAL.
FT  CDSIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT  PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT  ACT_SITE 299 299 PROBABLE.
SQ  SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150A6A CRC64;

Query Match 5.0%; Score 88; DB 1; Length 1849;
Best Local Similarity 19.0%; Pred. No. 49;
Matches 73; Conservative 47; Mismatches 128; Indels 136; Gaps 20;

QY 35 YTGPISTPVLTTPGD---RIIVDTIRDAFE-GAINSE-ODIPSOLLAMPFLNPONGPI 88
DB 17 YALTPTTEALVADDOVDYQIFRDAENKGFSGATNVEVRDKKNSL-----GSA 67

QY 89 MVNCAEKGDVLAIVIESMLPRGVDPV-----GICAMTPHFGGLTGDTLAMLNDP 138
DB 68 LPRGIPMIDFSDVDVKRIATLVNQQVGVGVKIVSGVSEL--HFNLSGN-----MNG 120

QY 139 LPEKVRMIKLISEKVYVSKRHLP-----YKPHIGTSLVSPET 176
DB 121 NAKSHRDVSEENRYITVEKNNEFTENVTSTTKKEQDAQRREDYVMPRLDKFT--EV 178

QY 177 DSINSLTPDNHGGNMVDPDIPGSIYTP-LVIRAPGRLFIGDAHACQGGGICGTAVET 235
DB 179 APTASTANNKGEYNNSD-----KYPAPVRLGSGSQFIYKKS-----RYQ 220

QY 236 SITTIK---VDLIKNNLSWPRME-----NAENIMTSGSARPLE---D 272
DB 221 LILTEKDKQGNLURNWVDGGDNLVGNAYTYGIAGTPYKVNHNGLIGFGNSKEHSD 280

```



```

FT DISULFID 278 465 BY SIMILARITY.
SQ SEQUENCE 465 AA: 53738 MW: E779B6014EC6FF68 CPC64:

Query Match      4.98; Score 86; LH 1; Length 455;
Best Local Similarity 24.8%; Pred. No. 12;
Matches 40; Conservative 20; Mismatches 5; Indels 50; Gaps 10;

QY 177 DSINSLTPINHGGNM---DVPDIG-----PGSIYPLVR--APGG-RIFIGDAHA 220
Db 194 DEAEWTKSFQGTNLFVAVHGLGSLRHSNNPKSIMYPTTYRLHPNFRLSAPDIRS 253
QY 221 CQG-----DGEICGTAVEPASITTI--KVCLIKNWQLSWPRNEN--AE 259
Db 254 IQSLYGAPVKNPISLMPGSPPTVCHQSLSFDVATVGDKIFFFKDW--FWWR-LFGSPAT 312
QY 260 NIMSIGSARP-----LEDATRIA-----YRDLIYWLVED 258
Db 313 NIISISSMWP:IPSGIOAAVEIGGRNQJLFPDEKYWLINN 353

```

Search completed: July 3, 2003, 15:47:46  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compaq Inc.

OM protein - protein search, using sw model

Run on: July 3, 2003, 15:44:52 : Search time 80 Seconds  
(without alignments)  
844.793 Million cell updates/sec

Title: US-10-086-082-2

Perfect score: 1748  
Sequence: 1 MKWDEESIMAKRGVGAGRKPR.....GNVDPKTYTGAMLNKLNLY 328

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL21:  
1: sp.archaea.\*  
2: sp.bacteria.\*  
3: sp.fungi.\*  
4: sp.human.\*  
5: sp.invertebrate.\*  
6: sp.mammal.\*  
7: sp.mhc.\*  
8: sp.ordanelle.\*  
9: sp.phage.\*  
10: sp.plant.\*  
11: sp.podent.\*  
12: sp.virus.\*  
13: sp.vertebrate.\*  
14: sp.unclassified.\*  
15: sp.virus.\*  
16: sp.bacteriap.\*  
17: sp.archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	368	21.1	317	16 Q92P43	Q92P43 rhizobium
2	363	20.8	300	16 Q9KGN3	Q9KGN3 bacillus
3	355.5	20.3	299	16 Q8R8S5	Q8R8S5 thermoanaer
4	348	19.9	344	17 Q8YIV6	Q8YIV6 pyrococcus
5	342	19.6	318	17 Q9TV26	Q9TV26 sulfobolus
6	332.5	19.0	377	17 Q9YEQ1	Q9YEQ1 aeropyrum p
7	327	18.7	318	17 Q976P0	Q976P0 sulfobolus
8	322.5	18.4	304	16 Q9A553	Q9A553 caulobacter
9	304.5	17.4	298	17 Q9V075	Q9V075 pyrococcus
10	300.5	17.2	285	16 Q9WXX3	Q9WXX3 thermotoga
11	269	15.4	328	16 Q8YIV6	Q8YIV6 thermoanaer
12	261.5	15.0	299	16 Q9Z7N6	Q9Z7N6 listeria in
13	239.5	13.7	411	3 Q9C453	Q9C453 emericella
14	233.5	13.4	410	3 Q9URV7	Q9URV7 schizosacch
15	233	13.3	296	16 Q8Y452	Q8Y452 listeria mo
16	208	11.9	432	10 Q9SZB9	Q9SZB9 arabidopsis

17	198	11.3	454	10 Q34DX1	Q94dx1 oryza sativ
18	133.5	11.1	435	16 Q8RCU9	Q8rcu9 thermoanaer
19	166	10.6	452	10 Q94CA8	Q94ca8 arabidopsis
20	173	10.2	389	17 Q587S5	Q587S5 pyrococcus
21	167.5	9.6	382	10 Q94AP4	Q94ap4 arabidopsis
22	130	7.4	307	10 Q9SZF0	Q9szf0 arabidopsis
23	107	6.1	238	16 Q8Y6A5	Q8y6a5 listeria m
24	105	6.0	842	10 Q9FLJ8	Q9flj8 arabidopsis
25	104	5.9	344	3 Q96UW2	Q96uw2 trichoderma
26	102	5.8	238	16 Q92AL8	Q92al8 listeria in
27	100.5	5.7	471	3 Q9P634	Q9p634 neurospora
28	99.5	5.7	561	17 Q8U4J7	Q8u4j7 pyrococcus
29	99	5.6	977	16 Q9ENV6	Q9env6 rhizobium
30	98.5	5.6	348	16 Q9JUS4	Q9jus4 neisseria m
31	98	5.6	996	4 Q8TD25	Q8td25 homo sapien
32	97.5	5.6	949	16 Q9JZP1	Q9jzpl neisseria m
33	97	5.5	380	3 Q9V772	Q9v772 emericella
34	96.5	5.5	432	16 Q97MG8	Q97mg8 clostridium
35	96	5.5	945	17 Q8ZXT6	Q8zxt6 pyrobaculum
36	94.5	5.4	470	16 Q8RZM1	Q8rzm1 anabaena sp
37	94.5	5.4	519	16 Q8R8R6	Q8r8r6 thermoanaer
38	94.5	5.4	613	16 Q92L36	Q92l36 rhizobium m
39	94	5.4	344	3 Q96VR2	Q96vr2 trichoderma
40	94	5.4	655	16 Q9RYM2	Q9rym2 deinococcus
41	93.5	5.3	466	16 Q8XNM0	Q8xnm0 clostridium
42	93	5.3	682	3 Q9P8S3	Q9p8s3 aspergillus
43	92.5	5.3	505	17 Q9V2G8	Q9v2g8 pyrococcus
44	92.5	5.3	513	17 Q8TUQ7	Q8tuq7 methanosarc
45	92	5.3	547	2 Q9E111	Q9eti1 citrobacter

#### ALIGNMENTS

RESULT :  
Q92P43 PRELIMINARY: PRT: 317 AA.  
UT Q92P43: C-DEC-2001 (TREMREL: 19, Created)  
UT 01-DEC-2001 (TREMREL: 19, last sequence update)  
DT 01-MAR-2002 (TREMREL: 20, last annotation update)  
DE Hypothetical protein R01950.  
GN R01950 OR SMC04290.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OZ Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID:382;  
RN 1..1..  
RP SEQUENCE FROM N.A.  
RC STRAIN:1021;  
PK MEMPHINE:21396507: PubMed:11481430;  
KA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
Rostard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
Gedric V., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
Pohl T., Portetelle D., Puelher A., Purnelle B., Ramsperger U.,  
Renard C., Thebault P., Vandenbol M., Weidner S., Gallbert F.,  
RA "Analysis of the chromosome sequence of the legume symbiont  
PT Sinorhizobium meliloti strain 1021.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL: AL591789: CAC46529.1: -;  
DR InterPro: IPR004304: FmdA\_AmdA.  
DR Pfam: PF03069: FmdA\_AmdA: 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 317 AA: 33877 MW: 7705D716F6B9EE2A CRC64;

Query Match 21.1%; Score 368; DB 16; Length 317;  
Best Local Similarity 32.4%; Pred. No. 9.2e-24;  
Matches 102; Conservative 43; Mismatches 128; Indels 42; Gaps 10;  
Qy 23 HLLTEEMOKFHYTIGYSTPVLTIPEGDRITVTRDAFEAGINSEQDIPS-QLIKMPL 81  
Db 16 HH-----HYGWBHSIPPTVAPGSRLEFNCNCLDSGSGHFTADSTVADYSTVDFKV 66









```

RC STRAIN-ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RL structure and evolution.";
RL Submitted (20L-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248285; CAP49430.1; -
DR InterPro: IPR004304; FmdA_AmdA;
DR Pfam: PF03069; FmdA_AmdA;
DR Hypothetical protein: Complete proteome.
KW SEQUENCE 298 AA; 32456 MW; 47CF8AD3C3362E22 CRC64;

Query Match 17.4%; Score 304.5; DB 17; Length 299;
Best Local Similarity 30.7%; Pred. No. 2.8e-18;
Matches 94; Conservative 45; Mismatches 120; Indels 47; Gaps 10;

QY 35 YTGIPSTVLTIEPDRIIVDTRAFEGAINSEQDIPSQLKMPF--LNPONGPIMVNG 92
DB 14 YSGPNKKEVARAKGEIVFOTLDALGQVKSFE--T:EKIDFSRVNPATGPLYVEG 70
QY 93 AEGSOVLAVYIESMLPRGVDYXGICAMIPHFQGLTGTDITAMINDPLPEKV-----RMK 147
DB 71 AKRGGLRLVDLIDIKVEG-----KGAVVTAPGAGVGGKVKVERPOTRICE 114
QY 148 LDESEKYVWSKRHTLPYKPHIGTILSVSPEDSINSLTPDNHGGNMDVDPGSGSITYPLVR 207
DB 115 VKGQVIF-KGKIPAMPKIGVIGVAYD-EEVPTGPGKHGGNMDINLRKGTTFYFVF 172
QY 208 APGRFLFGDAHACOGDGEICGTAVEFAFASITIKVDLKNQLSWPRMENAINSIGRA 267
DB 173 VDGAYLAIGDLHVMGGEVGSACEVSGEV-VRVTPNEG-KLENPLLTIESDFVLVSD 231
QY 268 RPLDATRAYRLIYWLVEDEGFE-----QWD-AYMLISQCGKVLGNMVDPKIV 318
DB 232 ENLQKAEAV-----SLGVEALRKSNDISMDFAFMASLVNDVEISLVDPKIV 282
QY 319 GAHLNK 324
DB 263 RVRIKX 288

RESULT 10
Q9WXX3 PRELIMINARY; PRI: 285 AA.
AC Q9WXX3:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
LT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Acetamidase, putative.
GN TM0119.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotocaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Neider W.C., Ketchum K.A.,
RA McDonald L., Otterback T.R., Malek J.A., Liuher K.D., Garrett M.K.,
RA Stewart A.M., Sutton M.D., Pratt M.S., Phillips C.A., Richardson B.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White G.,
RA Salzberg S.L., Smith H.G., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima.";
DR EMBL: AE001698; AAD35213.1; -
DR TIGR: TM0119; -
DR Nature 399:323-329(1999)
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA; 1.
KW Complete proteome.
KW SEQUENCE 285 AA; 31196 MW; 131E9167BD6DDDBE CRC64;

Query Match 17.2%; Score 300.5; DB 16; Length 285;

```

```

Best Local Similarity 29.7%; Pred. No. 5.8e-18;
Matches 85; Conservative 53; Mismatches 125; Indels 33; Gaps 9;

QY 35 YTGIPSTVLTIEPDRIIVDTRAFEGAINSEQDIPSQLKMPF--LNPONGPIMVNG 92
DB 11 YSFSANAPVFEYVPGGVVFTLDALGGSYD-----KIDFSKYNPATGPFVNG 60
QY 93 AEGSOVLAVYIESM-LPRGVDYXGICAMIPHFQGLTGTDITAMINDPLPEKVMKILDE 151
DB 6: VKGQDYLKVRKRIELPR---RGMIVTSGFG-----VLGDEV-EGEHTKELEIE 106
QY 152 KYVWSKRH---TLPYKPHIGTILSVSPEDSINSLTPDNHGGNMDVDPGSGSITYPLVRA 208
DB 107 K--WAVLFDGVRIPLEPMVGIVGAPQEGEYPTGTAHRHGGNMIDKEITENVTVHLPVFO 164
QY 209 PGGRFLFGDAHACOGDGEICGTAVEFAFASITIKVDLKNQLSWPRMENAINSIGRA 268
DB 165 EGALLAIGDGHATMGDEVCVSACEVPAKVVEIDVSK-EIKWPVETNDYIIVSLP 223
QY 269 PLEDATRAYRLIYWLVEDEGFEQWDAYMLISQCGKVLGNMVDPKIVTGAMLNK 324
DB 224 DTEALKVENTREIVMLTORRKTIPP'DAYMIASLSVDGISOLVNPAPKAKARIPK 279

RESULT 11
Q8YVY6 PRELIMINARY; PRI: 328 AA.
AC Q8YVY6:
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 02-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein: All0706.
GN ALL0706.
OS Arabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
KT cyanobacterium Arabaena sp. strain PCC 7120.";
RC DNA Res. 8:205-213(2001);
DR EMBL: AF003583; BAB72663.1; -
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA; 1.
KW Hypothetical protein: Complete proteome.
KW SEQUENCE 328 AA; 35627 MW; DE5FBA0E4AF4165D CRC64;

Query Match 15.4%; Score 269; DB 16; Length 328;
Best Local Similarity 27.6%; Pred. No. 3.8e-15;
Matches 91; Conservative 59; Mismatches 140; Indels 40; Gaps 9;

QY 21 VTHLTEMOKEPHYTIGPVS---TPVLIEPDRIIVDTRAFEGAINSEQDIPSQLK 77
DB 1 MTHILKATKSTVE--IGGFSHLJEPALKVDSST--DVETITGY-----YIDKAPPEFLT 54
QY 78 MPFIN-----PQ-----NGPTVMGAFKGDVLAVYIESMLPRGVDYXGICAMIP 121
DB 55 PEFIDICQLPPEPKIAAGPHLLTJGPIYVKDAEPGDVLEVKLEAIAPS--LPVGNNAIRS 112
QY 122 HFGGLTGTDITAMINDPLPEKVMKIL-----DSEKYVWSKRHTLPYKPHIGTILSVSPED 177
DB 113 GWG-----ALPNOFTOPALRFILNLANNTAEPPNSGKIKPLTLPFGILGVATPEN 164
QY 176 SINSLTPDNHGGNMDVDPGSGSITYPLVRAPGRFLFGDAHACOGDGEICGTAVEFAFI 237
DB 165 ARSSVPPGYGGNIDNRELQAGSRIFLIPVPGGLFSLGDSGSAQGDGEVNTALETSMN 224

```

```
QY 238 TTIKVLKKNQLSWPRMENENIMSGSRPPLEDATRAIYRDLIYMWVEDFGFEQDAY 297
DB 225 GRIOLTLRKHJHFTAPIAETPTDITITGFAPTDAALEQALKNMIDFLRFVNLSPEDAY 284
QY 298 MLLSQGKVRIGNVMD-PKYIVGAMLNKE 326
DB 285 VLCSLVNFRITOVVNSPKNGVHGILLPKL 314

RESULT 12
Q927N6
ID Q927N6 PRELIMINARY: PRT: 299 AA.
AC Q927N6
DT 01-DEC-2001 (TRENBLrel, 19, Created)
DI 01-DEC-2001 (TRENBLrel, 19, Last sequence update)
DI 01-JUN-2002 (TRENBLrel, 21, Last annotation update)
DE Hypothetical protein lin2752.
GN LIN2752.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Franquet L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Blocker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Echoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurgey C.,
RA Etian K.-D., Fstai H., Garcia-Jel Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Bain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kust F., Kurapat G.,
RA Madueno E., Mailounar A., Marz Vicente J., Ng E., Nodjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Miaz J.-C., Porcelli K.,
RA Remmel B., Rose M., Schluerer T., Simoes N., Tietz P.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001)
DR EMBL: AL596173; CAC97978.1;
DR Lifestlist: LIN02752;
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 299 AA: 33:83 MW: 879878A0B90A70D9 CRC64;
Query Match 15.0%; Score 261.5; DB 16; Length 299;
Best Local Similarity 29.0%; Pred. No. 1.5e-14;
Matches 79; Conservative 44; Mismatches 112; Indels 37; Gaps 7;

QY 43 PVLTIEPGDRIIVTDRAFEAGINSEQDIPSQLKMPFLNPNQPTMVNGAEKGDVIAY 102
DB 20 PVLTIVKDSVVTIKIHFNQINKEQLHYGEIDMKQF-SPTIGPICIEAREFGDILLAV 78
QY 103 TESMLPRGVDPYGCAMIPHFEGLTGTDLATMLNDP-----LPEK-VRMKILDSEKV 153
DB 79 IEKJT-----LTSKDV-VLLNPGTIGVTDLLPNNCIRRYKIKONKI 219
QY 154 YNSKRTILPYKPHGTLSVSPEDISINSL-----TPDNHNGMNVDPDIPGSIYFAYRAP 209
DB 120 IYSDIEHVQLOKTIIGL-----KTESIHQTFPHPIKGGTGLSPDITEGATIFIVERK 175
QY 210 GGRFLIGDAHACQGDGEI--CGTAVEFASITITKVDLIKNNQLSWPRMENENIMSGSAR 269
DB 176 GAMLHVAIDRATIGTKITTSIAEVPREVIIRLQGLKWK-APTPII-HKNNLICLSFLT 235
QY 270 LEDATRIAYRDLIYMWVEDFGFEQDAYMLLS 301
DB 236 TEKATKKALQNNMLNVMESDKITLEDA-FLLS 267

RESULT 13
Q9C453
ID Q9C453 PRELIMINARY: PRT: 411 AA.
AC Q9C453
DT 01-JUN-2001 (TRENBLrel, 17, Created)
DI 01-JUN-2001 (TRENBLrel, 17, Last sequence update)
DI 01-JUN-2002 (TRENBLrel, 21, Last annotation update)
DE Formamidase (EC 3.5.1.49).
GN FMUS.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX NECLINE=21066161; PubMed=11139495;
RA Fraser J.A., Davis M.A., Hynes M.J.;
RT "The Formamidase Gene of Aspergillus nidulans: Regulation by Nitrogen
RT Metabolite Repression and Transcriptional Interference by an
RT Overlapping Upstream Gene.";
KL Genotics 157:119-131(2001).
DR EMBL: AF274009; AAG60585.1;
DR InterPro: IPR004304; FmdA_AmdA.
DR Pfam: PF03069; FmdA_AmdA; 1.
KW Hydrolyase.
SQ SEQUENCE 411 AA: 44907 MW: D36B46A5243AF242 CRC64;
Query Match 13.7%; Score 239.5; DB 3; Length 411;
Best Local Similarity 25.5%; Pred. No. 1.9e-12;
Matches 95; Conservative 45; Mismatches 123; Indels 109; Gaps 17;

QY 30 OKFPHYTIGTYSTVLTIEPGDRIIVTDRAFEAGI---NSEQDIPS-QLLKMPFLNPNQ 85
DB 21 QKGLHNRHRRPDPSPSYGTIANNEVVKISCLDWTGGQIKNDDSAADDIKNVDITQIHYL---S 77
QY 85 GPIWNGAEKGDVIAYTESMLPRGVDPYGCAMIPHFEGLTGTDLATMLNDPPEKVRM 145
DB 78 GPFDTIAREPGDVLVLIQCVQPFEDQPFQVFARENG-----GGFDEIYPPAKA 131
QY 146 TKLDSEKVVYWSKRITLPYKPHI-----GTLVSVP-----EIDISINSLIP 184
DB 132 I-WDFEGIFGSSRH-----PHVREAGLIHPGLGCAPSAEVLAFNHRREGELIAANTLGR 186
QY 185 D-----NUGGNMVDVFDIGPSIYFAYRAP 209
DB 187 EVAKPPSPNVHAGSADGELAATIGRGARTIPGRPEHNGCNDIKNI-SRGSKVLVPHVP 246
QY 210 GGRFLIGDAHACQGDGEI--CGTAVEFASITITKVDLIKNNQLSWPRMENENIMSGS 257
DB 247 GAKFSVGLHFESQGDGEISFCG-AIEMAGVITLKTFTVKD---GMAKMAMKSPFIHPGPV 302
QY 268 RP-----LEDATRIAYRDLIYMWVE---DFGFEQDAYMLLSQ 302
DB 303 EPQFGPGRYLTPEGFSVDEKQKQHYLDAT-VAYKQTCRLVIEYLRRYGNDYQIYLLS- 360
QY 303 CCKVR--LGNMVD 313
DB 361 CAPVOGHIAGLVD 373

RESULT 14
Q9URY7
ID Q9URY7 PRELIMINARY: PRT: 410 AA.
AC Q9URY7
DT 01-MAY-2000 (TRENBLrel, 13, Created)
DI 01-MAY-2000 (TRENBLrel, 13, Last sequence update)
DI 01-DEC-2001 (TRENBLrel, 19, Last annotation update)
DE Formamidase-like protein.
GN SpAc869_04.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
```

